

# Journal of Experimental Criminology

## What can we do to reduce disciplinary school exclusion? A systematic review and meta-analysis

--Manuscript Draft--

<b>Manuscript Number:</b>	JOEX-D-18-00024R1	
<b>Full Title:</b>	What can we do to reduce disciplinary school exclusion? A systematic review and meta-analysis	
<b>Article Type:</b>	Original Research	
<b>Keywords:</b>	School exclusion; school suspension; systematic review; meta-analysis	
<b>Corresponding Author:</b>	Sara Valdebenito University of Cambridge UNITED KINGDOM	
<b>Corresponding Author Secondary Information:</b>		
<b>Corresponding Author's Institution:</b>	University of Cambridge	
<b>Corresponding Author's Secondary Institution:</b>		
<b>First Author:</b>	Sara Valdebenito	
<b>First Author Secondary Information:</b>		
<b>Order of Authors:</b>	Sara Valdebenito Manuel Eisner David Farrington Maria Ttofi Alex Sutherland	
<b>Order of Authors Secondary Information:</b>		
<b>Funding Information:</b>	Nuffield Foundation (Grant EDU/41720) BECA Chile	Dr. Manuel Eisner Dr. Sara Valdebenito
<b>Abstract:</b>	<p><b>Objectives:</b> To systematically review and quantitatively synthesize the evidence for the impact of different types of school-based interventions on the reduction of school exclusion or suspension.</p> <p><b>Methods:</b> A systematic search of 27 databases including published and unpublished literature was carried out between September and December 2015. Eligible studies evaluated interventions intended to reduce the rates of exclusion; targeted children from ages four to 18 in mainstream schools; and reported results of interventions delivered from 1980 onwards. Only randomised controlled trials were included. Two independent reviewers determined study eligibility, extracted data and rated the methodological quality of studies.</p> <p><b>Results:</b> Based on the thirty-seven studies eligible for meta-analysis, under a random effects model, results showed that school-based interventions significantly reduced school exclusion during the first six months after implementation <math>SMD=.30</math>; (95% CI .20 to .41; <math>p&lt;.001</math>). The impact at follow-up (i.e., 12 or more months) was reduced by half and it was not statistically significant. Heterogeneity was mainly explained by the role of the evaluator: independent evaluators reported lower effect sizes than researchers involved in the design and/or delivery of the intervention. Four approaches presented promising and significant results in reducing exclusion: enhancement of academic skills, counselling, mentoring/monitoring, and skills training for teachers.</p> <p><b>Conclusions:</b> Results suggest that school-based interventions can be effective in reducing school exclusion in the short term. Some specific types of interventions show more promising and stable results, but, based on the small number of studies involved</p>	

	in our calculations, we suggest that results are interpreted with caution.
<b>Response to Reviewers:</b>	Responses have been attached (File:Revisions_rev)

**TITLE: What can we do to reduce disciplinary school exclusion? A systematic review and meta-analysis.**

**Sara Valdebenito<sup>a</sup>, Manuel Eisner<sup>a</sup>, David P. Farrington<sup>a</sup>, Maria M. Ttofi<sup>a</sup> and Alex Sutherland<sup>b</sup>.**

- a. Institute of Criminology, University of Cambridge
- b. RAND Europe

**Corresponding author:**

Sara Valdebenito Email: [sv331@cam.ac.uk](mailto:sv331@cam.ac.uk); [sara.valdebenito@gmail.com](mailto:sara.valdebenito@gmail.com). Telephone: +44 -1223767373

**Declaration of interest:**

None of the authors have any financial interests related to findings reported in this review.

**Sources of support:**

The Nuffield Foundation is an endowed charitable trust that aims to improve social well-being in the widest sense. It funds research and innovation in education and social policy and also works to build capacity in education, science and social science research. The Nuffield Foundation has funded this project, but the views expressed are those of the authors and not necessarily those of the Foundation. More information is available at [www.nuffieldfoundation.org](http://www.nuffieldfoundation.org)

**Acknowledgements:**

Special thanks to Mr. Aiden Cope who assisted the team with information retrieval and coding.

**Manuel Eisner** is Wolfson Professor of Criminology and Director of the Violence Research Centre at the Institute of Criminology at the University of Cambridge. He is also Professor of Developmental Criminology at the University of Zurich. His research interests include macro-level comparative violence research, micro-level predictors of violence, violence prevention, child and adolescent development, aggression and anti-social behaviour.

**Sara Valdebenito** is a Research Associate at the Violence Research Centre at the University of Cambridge. She received her doctoral degree in Criminology from the University of Cambridge. Her major research interests include violence prevention, evaluation, child maltreatment, adolescence, anti-social behaviour, and early childhood development.

**David P. Farrington** is Emeritus Professor of Psychological Criminology at the Institute of Criminology, Cambridge University. He has been chosen to receive the John Paul Scott Award of the International Society for Research on Aggression in 2018, for significant lifetime contributions to aggression research, and the Herbert Bloch Award of the American Society of Criminology in 2018, for outstanding service contributions to criminology. He received the Stockholm Prize in Criminology in 2013. His major research is in developmental criminology, and he is Director of the Cambridge Study in Delinquent Development, which is a prospective longitudinal survey of over 400 London males from age 8 to age 61. In addition to 750 published journal articles and book chapters on criminological and psychological topics, he has published 108 books, monographs and government publications, and 152 shorter publications (total = 1,010).

**Maria M. Ttofi is the University Lecturer in Psychological Criminology at Cambridge University, UK.** She is interested in the development of antisocial behaviour, crime, and violence through the life course. Earlier research has focused on highly aggressive and victimized children; the long-term impact of negative childhood experiences on mental health and offending behaviour in adult life; and early prevention and intervention research against youth aggression, antisocial behaviour and victimization. Current research looks at the interplay between mental health and crime and in effective strategies for healthy reintegration of antisocial individuals in society.

**Alex Sutherland** is a Senior Research Leader at RAND Europe. His areas of research are criminal justice and education, and he leads RAND's trials research, including several large-scale RCTs in education.

## ABSTRACT

**Objectives:** To systematically review and quantitatively synthesise the evidence for the impact of different types of school-based interventions on the reduction of school exclusion.

**Methods:** A systematic search of 27 databases including published and unpublished literature was carried out between September and December 2015. Eligible studies evaluated interventions intended to reduce the rates of exclusion; targeted children from ages four to 18 in mainstream schools; and reported results of interventions delivered from 1980 onwards. Only randomised controlled trials were included. Two independent reviewers determined study eligibility, extracted data and rated the methodological quality of studies.

**Results:** Based on the thirty-seven studies eligible for meta-analysis, under a random effects model, results showed that school-based interventions significantly reduced school exclusion during the first six months after implementation  $SMD=.30$ , 95%  $CI [.20, .41]$ ,  $p<.001$ . The impact at follow-up (i.e., 12 or more months) was reduced by half and it was not statistically significant. Heterogeneity was mainly explained by the role of the evaluator: independent evaluators reported lower effect sizes than researchers involved in the design and/or delivery of the intervention. Four approaches presented promising and significant results in reducing exclusion: enhancement of academic skills, counselling, mentoring/monitoring, and skills training for teachers.

**Conclusions:** Results suggest that school-based interventions can be effective in reducing school exclusion in the short term. Some specific types of interventions show more promising and stable results, but, based on the small number of studies involved in our calculations, we suggest that results are interpreted with caution.

**Keyword:** School exclusion, school suspension, systematic review, meta-analysis

## INTRODUCTION

Schools use different procedures to manage disruptive behaviour. Both punitive (e.g., loss of privileges, additional homework or detention hours) and non-punitive strategies (e.g., targeted behavioural support for at-risk students or interventions to reduce violence) are aimed at keeping order in schools. Among punitive responses, school exclusion is normally seen as one of the most serious sanctions. School exclusion can broadly be defined as a disciplinary measure imposed in reaction to students' misbehaviour (e.g., violations of school policies or laws) by a responsible authority. Exclusion entails removing the pupil from regular teaching for a period, during which he or she is not allowed to be present in classrooms (in-school exclusion) or on school premises (out-of-school exclusion). Fixed-term exclusions consist of a prescribed number of hours or days (Cornell et al. 2011), whereas permanent exclusion involves the pupil being transferred to a different school, or educated outside of the regular education system (Spink 2011).

In England fixed-term exclusion affects 4.3% of the school population, which corresponds to around 1,590 fixed term exclusions per day. The percentage of permanent exclusion applies to only 0.08% of the school population, which can be translated into 35.2 permanent exclusions per day on average. The national figures suggest that students in secondary-level education (8.4% of the school population) as well as those in special education (12.5%) are the most likely to experience permanent and fixed term exclusion (Department for Education [DfE], 2017). In the United States, the most recent data provided by the Department of Education concluded that 7.4% (3.5 million) of students were suspended in school, 7% (3.4 million) were suspended out-of-school, and less than 1% were subject to expulsion (around 130,000 students). Black students and those presenting disabilities are, respectively, three and two times more likely to be excluded compared to White and non-disabled pupils (U.S. Department of Education, 2012).

Data suggest that exclusion is largely a male experience (e.g., DfE, 2013, 2017; Liu 2013; Ministry of Education Ontario Canada 2014), disproportionately affecting adolescents from economically disadvantaged families (Hemphill et al. 2010; Nichols 2004; Skiba et al. 2012) as well as those from ethnic minorities (DfE 2012; Noltemeyer and McLoughlin 2010; Skiba et al. 2011). Notably, recent multivariate analysis points out that racial disproportionality in exclusion still remains significant after controlling for behaviour, number and type of school offences, age, gender, teacher's ethnicity, and socio-economic status (e.g., Fabelo et al. 2011; Noltemeyer and McLoughlin 2010; Rocque and Paternoster 2011; Skiba et al. 2002). Other pieces of research have specified the role of Special Education Needs (SEN) in the rates of exclusion (Achilles et al. 2007; Sullivan et al. 2014). Research by Bowman-Perrott et al. (2013) specifically concluded that children identified

with emotional/behavioural disorders (OR=3.95,  $p<.05$ ) attention-deficit/hyperactivity disorders (OR=4.96,  $p<.05$ ) and learning disabilities (OR=2.54,  $p<.05$ ) were more likely to be suspended or expelled from school, than their non-affected peers.

Some literature looking at the correlation between exclusionary punishments and behaviour suggests that such punishments could prompt a spiral into more challenging behaviour (American Psychological Association Zero Tolerance Task Force [APA ZTTF] 2008; Chin et al. 2012; Sharkey and Fenning 2012). In fact, the use of disciplinary exclusion has been correlated with serious behavioural problems such as antisocial behaviour, delinquency and early entry into the juvenile justice system (Costenbader and Markson 1998; Hemphill and Hargreaves 2010). However, particularly in relation to behavioural problems later in life, it is not clear whether exclusion is merely a marker of underlying problems that manifest in contact with other institutions such as the police, juvenile justice and prisons.

Evidence also suggests that periods of exclusion may have detrimental effects on pupils' learning outcomes. Exclusion is accompanied by missed academic activities, alienation, demotivation regarding academic goals (Arcia 2006; Brown 2007; Michail 2011) low marks and drop-out (Noltemeyer and Ward 2015). In the long-term, opportunities for training and employment seem to be considerably reduced for those who have repeatedly been excluded (Brookes et al. 2007; Massey 2011).

Different programmes have attempted to reduce the prevalence of exclusion. Although some of them have shown promising results, so far, no comprehensive systematic review has examined the overall effectiveness of them. The main goal of the present research is to systematically examine the available evidence for the effectiveness of different types of school-based interventions aimed at reducing disciplinary school exclusion. Secondary goals include comparing different approaches and identifying those that could potentially demonstrate larger and more significant effects.

The research questions underlying this project are as follows:

- Do school-based programmes reduce the use of exclusionary sanctions in schools?
- Are some school-based approaches more effective than others in reducing exclusionary sanctions?
- Do participants' characteristics affect the impact of school-based programmes on exclusionary sanctions in schools?

- Do characteristics of the interventions, implementation, and methodology affect the impact of school-based programmes on exclusionary sanctions in schools?

## METHODS

The title of the present review was registered in The Campbell Collaboration Library of Systematic Reviews on January 2015. The review protocol was approved in November 2015. The full report is available at (blinded).

### Inclusion criteria

*Participants.* The present review included students aged from 4 to 18 years irrespective of nationality, ethnicity, language, and cultural or socio-economic background. Reports involving students who presented special education needs, but were educated in mainstream schools, were included in this review. However, reports involving students with serious mental disabilities or those in need of special schools were excluded. Therefore, the results of this review are intended to be generalisable to mainstream populations of students in non-specialised schools from all the included countries.

*Included interventions.* We included interventions that were defined as school-based: that is, delivered on school premises, or supported by schools with at least one component implemented in the school setting. Interventions in the present review cover a wide range of psychosocial strategies for targeting students (e.g., Cook et al. 2014), teachers (e.g., Ialongo et al. 2001) or the whole school (Bradshaw et al. 2012)

*Study design.* The present review only includes randomised controlled trials. This type of study has the strengths of controlling for measured and unmeasured variables, thereby eliminating selection effects (Shadish et al. 2002; Sherman et al. 2002).

*Outcomes.* Eligible studies addressed school exclusion as an outcome. School exclusion was defined as an official disciplinary sanction imposed by an authority and consisting of the removal of a child from their normal schooling. We included studies testing fixed or permanent, as well as in-school and out-of-school, exclusion. For any identified study that reported findings on school exclusion, we also coded the effects of the intervention on internalising and externalising problem behaviour (Achenbach 1978; Achenbach and Edelbrock 1979; Farrington 1989).



*Included literature.* Databases and journals were searched from 1980 onwards, with the aim of including more contemporary interventions. To minimise publication bias, we included both published and unpublished reports.

### **Search strategy**

Searches were run in 27 different electronic databases between September and December 2015 (see Table 1). For each database, we ran pilot searches using a set of key terms included in Table 2.

#### *TABLE 1 AROUND HERE*

As planned, we contacted key authors requesting information on primary studies that could potentially be integrated in this review. Lists of references from previous primary studies and reviews related to the intervention/outcomes were appraised (e.g., Burrell et al. 2003; Mytton et al. 2009). There was no language restriction placed on the searches, as long as the abstract was written in English.

#### *TABLE 2 AROUND HERE*

### **Data extraction and risk of bias appraisal**

Two independent coders (AS & AC) extracted data from each eligible study (i.e., descriptive data and data for effect size calculation). Discrepancies during coding were moderated by agreement, but when the information reported was contentious, senior members of the team (ME & DF) were consulted for further input. When the data for calculation of an effect size was incomplete, we tried to find more details in other sources (e.g., published protocols or reports) or the lead researcher was contacted regarding the additional data needed. Endnote X7 software was used to manage references, citations and documents. Descriptive data extracted to characterise studies was inputted in Stata v.13 to produce inferential/descriptive statistics. To check code consistency across studies, we used Cohen's Kappa coefficient (Cohen 1960).

The methodological quality of each included study was evaluated using the Cochrane Effective Practice and Organisation of Care (EPOC) risk of bias tool (Cochrane Effective Practice and Organisation of Care 2007). The instrument was developed to assess the risk of bias for primary studies (i.e., internal validity) and is part of the guidelines provided by the Cochrane Collaboration. EPOC evaluates eight domains using a 3-point scale

(i.e., low risk, high risk, unclear risk). Two coders (SV & AS) independently applied the EPOC tool to each study at different locations and resolved conflicts in a final review meeting.

## Data analysis

Standardised Mean Differences (SMD) were calculated to measure the treatment effect along with 95% confidence intervals. For identifying heterogeneity in our results, we used Tau-squared, the *Q*-statistic and  $I^2$ . To explore sources of heterogeneity, we included moderator analysis involving models analogous to ANOVA as well as meta-regression. Effect sizes were coded such that a positive effect reflects the outcomes favouring the treatment group. Meta-analyses, sub-group analyses and meta regression were performed using the Comprehensive Meta-Analysis software (CMA, Version 3).

Where studies reported repeated measures of the outcome such as baseline and post treatment (e.g., Hawkins et al. 1988), SMDs were estimated as a synthesis index of the difference, representing the change between those different time points  $\mu_1$  and  $\mu_2$ . Because measures at baseline and post-treatment are positively correlated, we corrected the variance calculation as shown in equation 1 (Borenstein et al., 2009). Since the covariance statistic was not usually reported in primary research (and this was commonly the case in our set of included studies) we assumed a value equal to .75 (following Farrington and Ttofi 2009).

$$SE_{\mu_1 - \mu_2} = \sqrt{SE_1^2 + SE_2^2 - 2SE_1 SE_2 \sqrt{r}} \quad (1)$$

Since the inclusion of multiple time points would create statistical dependence due to the different measures based on the same subjects (i.e., correlated with each other), we calculate SMDs separately for studies reporting short-term (i.e., post-treatment) and long-term (i.e., follow-up) impact.

With cluster-randomised studies, standard errors were corrected as suggested by Borenstein et al. (2009) as well as Higgins and Green (2011). For the case of clustered data with dichotomous outcomes, the effective sample size was obtained by dividing the original sample size by the design effect. Once the *design effect* was identified (see equation 2), the square root of the design effect was multiplied by the original standard error of the log odds ratio.

$$1 + (k - 1) \times ICC \quad (2)$$

Since ICC was rarely reported in primary studies, we assumed a value of .05, based on the review of multiple

meta-analyses testing similar populations, produced by Ahn et al. (2012).

In the case of clustered studies with continuous outcomes measures, we followed the strategy suggested by Hedges (2007) and Spier et al. (2013). Effect sizes were computed using  $d_{T2}$  assuming equal cluster sample size:

$$d_{T2} = \left| \frac{\bar{M}_E - \bar{M}_C}{S_T} \right| \sqrt{1 - \frac{2(N-1)E}{E-2}} \quad (3)$$

In this equation  $\bar{M}_E$  and  $\bar{M}_C$  represent the overall means of the experimental and control groups and  $S_T$  is the total sample standard deviation estimated from the pooled sample standard deviation across the experimental and the control groups. Rho ( $\rho$ ) is the intraclass correlation.  $N$  is the total sample size and the sample size of the clusters is represented by  $n$ . Based on the characteristics of our data and following Spier et al. (2013) we assumed equal cluster size in our calculations.<sup>1</sup> When the clusters had different sizes we took a conservative approach, including the smallest cluster size in our calculation. The variance of the effect size was calculated by

$$d_{T2}^2 = \left( \frac{\bar{M}_E - \bar{M}_C}{S_T} \right)^2 (1 - (N-1)E) + \left( \frac{(N-2)(1-\rho)(1-\rho)^2 + N(N-2)\rho E^2 + 2(N-2)N\rho(1-\rho)}{2(E-2)[(N-2) - 2(E-1)N]} \right) \quad (4)$$

In this equation,  $N^E$  and  $N^C$  represent the experimental and control group sample sizes across clusters. As suggested by Higgins and Green (2011) in the event that the value of rho is not reported, analysts are advised to assume a reasonable value based in previous studies with similar populations (Ahn et al. 2012). As detailed previously, we have assumed a value of rho ( $\rho$ ) = .05.

The distribution of SMD effect sizes was examined to determine the presence of outliers. Following Lipsey and Wilson (2001), outliers were defined as those values which are more than two standard deviations from the overall mean of effect sizes. One outlier was detected (i.e., Collier 2002) and it was windsorised to the next closest value (Lipsey and Wilson 2001).

Since the present meta-analysis involved a wide range of decisions, we conducted sensitivity analysis to test the robustness of these decisions (Higgins and Green 2011). Specifically, we ran sensitivity analysis for the pre-post correlations (i.e., covariance) assumed to be .75. We re- ran the analysis using a correlation equal .50. We also ran sensitivity analysis testing the impact of the outlier and the impact of the windsorisation.

<sup>1</sup> The assumption of equal sample size seems to be a good approximation for the calculation of effect sizes. Hedges (2007) asserts that effect size calculations based on equal and unequal cluster sizes are not substantially different.

## RESULTS

### Results of systematic searches

As shown in Figure 1, different combination of search terms produced a total of 42,749 references from different electronic databases. Of these, 1,474 were kept because they were relevant to the study. After removing duplicates (N=958), a total of 516 unique manuscripts were saved for further assessment.

We next reviewed abstracts, methods and results sections. Four hundred and seventy one manuscripts were excluded because of the exclusion criteria displayed in Table 3 (for more details see Valdebenito et al. (2018)). As observed, the vast majority of reports (72%) were excluded because of the methodological design (i.e., they were not randomised controlled trials).

TABLE 3 ABOUT HERE

At the end, 37 papers presented enough statistical data for inclusion in our meta-analysis. Figure 1 shows a PRISMA flow diagram describing the flow of documents through the systematic screening and data-extraction process.

FIGURE 1 ABOUT HERE

After coding all includable papers, we calculated *Cohen's Kappa* for testing inter-rater reliability. In our review, *Cohen's Kappa* was equal to .76; *SD*=.81; *SE*=.036, reflecting a high level of agreement between coders.

### Characteristics of the included studies

We included studies presenting interventions carried out between 1980 and December 1, 2015 when we finished the searches (*Mdate*=2003; *SD*=9.5). Exceptionally the review involved three manuscripts published in 2016. In the first case, (Obsuth et al. 2016), the registered protocol of the study had been identified in our electronic searches and we were waiting for the final report. Sprague et al. (2016) was sent to us, because it included more information than a previous published version of the study published in previous years. The last study by Okonofua et al. (2016) was provided by an expert researcher in this field. As they were recent studies, matching our inclusion criteria, we decided to retain them in our analysis. No other study was sent to us after December 2015.

The included studies were roughly equal in terms of being published (51%) or unpublished reports (49%). All were written in English, and primarily presented studies from the United States (89.1%) and United Kingdom (8.1%). The remaining 2,8% represented 1 study where the country of the sample was not reported. Although we explored 27 databases, we were not able to find studies conducted in other locations.

Interestingly, only seven out of 19 (36.8%) experimental evaluations published in peer-reviewed journals disclosed a personal or organisational Conflict of Interest (CoI). In addition to the presence/absence of CoI statements, we evaluated studies on their potential conflict of financial interest (CoFI) by using a scale developed by Eisner and Humphreys (2012). It is a trichotomous scale that allow to identifies three levels of conflict: Unlikely (none of the study authors are programme developers or licence holders), Possible (a study author is a programme developer or collaborator with a programme developer AND the programme is not (yet) commercially available OR the business model is 'not-for-profit'), and Likely (study author is a programme developer or collaborator with a programme developer AND programme is commercially available AND business model is 'for-profit'). We found 18 studies (48.6%) where the CoFI was defined as 'unlikely' and 13 studies (35.1%) where we assessed a 'possible' CoFI and 4 studies where the information reported was not enough to judge CoFI. Finally, only two studies (5.4%) in our evaluation transparently declared information that allowed us to classify them as 'likely' to present a potential financial conflict of interest. In both cases the authors reported that one of the members of the evaluation team was related to the holder of a licensed programme that was evaluated.

The measures of exclusion reported in the included studies were mainly based on official records (81%) provided by schools or other official institutions. The average sample size was  $M_{size}=1,168$  ( $SD=3107.3$ ) participants. But this average should be cautiously interpreted since the included studies range from 20 to 13,498 participants (see Table 4 for further details).

### **Characteristics of participants in included studies**

In our meta-analysis, sampled students displayed a  $M_{age}=12.9$ ;  $SD= 2.8$ . Students attended schools with a high percentage of Black (54.1%) and Latino students (20.2%) on average. In fact, five of the included studies are based in schools where all pupils were Black (i.e., Barnes et al. 2003; Brett 1993; Collier 2002; Mack 2001; Reese et al. 1981). Interestingly, on average around two-thirds of pupils were eligible for/receiving free school

meals (FSM) ( $M_{FSM}=66.2\%$ ;  $SD=23.9\%$ ).

## **Intervention characteristics: targeted change, delivery and dosage**

The present review includes a wide range of school-based interventions. Broadly, 27% of the interventions were focused on changes at the school or teacher level, and 73% focused on changing pupils' skills/behaviours in order to affect exclusion rates. School staff (32.4%), or school staff assisted by external facilitators (24.3%), delivered the interventions. Most of those delivering the interventions were school psychologists/counsellors (32.2%) and, in two cases (5.4%), the intervention was delivered by police or probation officers.

Data on the role of the evaluator was coded. A high percentage of interventions were designed and/or delivered by the same researcher who evaluated the impact of the intervention (40.2%). On average, the included interventions lasted  $M=20$ ;  $SD=11.5$  weeks. 37.8% of the interventions lasted less than 12 weeks and an equal percentage were delivered over more than 24 weeks.

## **Research design and comparison condition**

Thirty-seven RCTs reporting 38 effect sizes were included in this meta-analysis. In 23 studies the control group received no-treatment (62.2%); six studies reported controls receiving 'treatment as usual' (16.2%), four experiments offered a placebo to the control group (10.8%) and four studies allocated controls via a waiting list (10.8%). Regarding the unit of randomisation, 70.3% of our studies randomised individuals, and almost 30% (11/37 studies) randomised clusters of students; either entire schools or classrooms (see Table 4 for further details).

### **1 Description of the interventions**

A wide range of school-based interventions were included in our review. For analytical purposes, we grouped them into nine categories. In some specific cases this grouping could be restrictive, because some of the interventions involve multiple components, but we have attempted to create an exhaustive list of interventions.

*Enhancement of academic skills.* We found two effect sizes targeting the enhancement of academic skills as a strategy to improve academic performance, increase motivation and promote more adaptive behaviour. Edmunds et al. (2012) tested an intervention to boost the academic progress of students in order to facilitate

1  
2 their future access to college while in the case of Cook et al. (2014), the intervention involved academic  
3 remediation plus social skills training.  
4

5  
6 *After-school programmes.* Two effect sizes come from interventions that offered students after-school activities.  
7  
8 The intervention tested by Dynarski et al. (2003) was more focused on academic support and recreational  
9 activities, while Hirsch et al. (2011) tested an after-school programme offering students support and paid  
10 apprenticeships.  
11  
12

13  
14  
15  
16 *Mentoring/monitoring programmes.* Five effect sizes reported the impact of interventions focused on  
17 mentoring/monitoring. These programmes involved structured and supportive relationships between a young  
18 person who presents academic, emotional or behavioural difficulties and a non-parental adult, their mentor or  
19 tutor. These adults (counsellors, teachers or members from the community) served as a role models providing  
20 support (e.g., Brett 1993; Wyman et al. 2010) supervising academic performance, providing advice or  
21 counselling, and assisting the students with academic tasks (e.g., Johnson 1983; Peck 2006; Reese et al. 1981).  
22  
23  
24  
25  
26  
27  
28  
29

30  
31 *Social skills training for students.* We found nine effects sizes representing the impact of social skills training  
32 for students. These programmes were based on social learning and cognitive behavioural theories (e.g.,  
33 Burcham 2002; Collier 2002; Harding 2011; Hostetler and Vondracek 1995; Shetgiri et al. 2011; Smith 2004)  
34 and their goal is to enhance individuals' socio-cognitive, socio-emotional, and behavioural skills to regulate  
35 maladaptive conducts. Social skills training programmes typically consist of a curriculum with focused training  
36 modules. Some more specific programmes target communication skills (e.g., Obsuth et al. 2016) or approaches  
37 to reducing stress (e.g., Barnes et al. 2003).  
38  
39  
40  
41  
42  
43  
44  
45

46  
47 *Skills training for teachers.* We found four independent interventions targeting teachers' skills. These involve  
48 training in facilitating mutual respect between teachers and students (e.g., Okonofua et al. 2016) as well as  
49 training to establish clear classroom rules (e.g., Hawkins et al. 1988). Skills for teachers also involve strategies  
50 for working in an alliance with parents to promote students' engagement in school activities (e.g., Ialongo et al.  
51 2001).  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

*School-wide interventions.* Six effect sizes represent comprehensive interventions targeting systemic changes across the whole school (Bradshaw et al. 2012; Cornell et al. 2012; Lewis et al. 2013; Snyder et al. 2010; Sprague et al. 2016; Ward and Gersten 2013). They involve pupils, teachers, parents and sometimes also the community where the school is based. These programmes aim to create positive environments, with clear rules that promote good behaviour, learning and safety.

*Violence reduction.* Three effect sizes came from violence reduction programmes. Although these interventions could be classified as skills training, we have isolated them because they are specifically targeted at increasing self-control and reducing violence (e.g., Feindler et al. 1984; Mack 2001). We also included anger management programmes encouraging peaceful responses to conflict (e.g., Farrell et al. 2010).

*Counselling and mental health interventions.* We included three effect sizes primarily focused on the provision of counselling in schools (e.g., Berlanga 2004; Tilghman 1988) and on a more specialised provision from community mental health services (i.e., Panayiotopoulos and Kerfoot 2004).

*Other interventions.* Four effect sizes were classified in this general category. They encompass a community services programme (Allen et al. 1997), a multicomponent programme (Arter 2005), a career awareness intervention (Bragdon 2010) and a programme focused on character-building education, promoting civic behaviour and national values (Crowder 2001).

## **Risk of bias**

The methodological quality of each publication included in the review was evaluated using the EPOC risk of bias tool. Overall, our assessment demonstrates that a high number of the included studies lacked enough information to judge risk of bias in all areas (see Figure 2). For instance, 59% of studies were not clear on how they allocated units to treatment and control groups (e.g., Arter 2005; Brett 1993; Feindler et al. 1984; Okonofua et al. 2016) and 54% of studies failed to report enough data to judge whether or not allocation was properly concealed (e.g., Allen et al. 1997; Bragdon 2010; Russell 2007; Tilghman 1988). Similarly, blinding to allocation (54%) and treatment/control group equivalence (50%) were other areas that rarely had enough detail to allow assessment (e.g., Arter 2005; Cook et al. 2014; Russell 2007).

FIGURE 2 AROUND HERE



1 In terms of incomplete data or attrition, 49% of the studies were evaluated as presenting a low risk of bias. Low  
2 risk cases involved i) those studies reporting zero attrition; ii) studies where attrition was represented by a small  
3 percentage of cases; iii) when the proportion of missingness was equivalent in the treatment and control groups;  
4 and iv) when the researcher reported attrition, analysed it and used methods to deal with attrition (e.g., multiple  
5 imputation, full information maximum likelihood or intention to treat rather than assessing the effect of  
6 treatment on the treated).

15 Among the eight criteria evaluated by EPOC, contamination or spill-over was probably the main threat to the  
16 validity of results among the RCTs included in our review. More than three-quarters of studies (76%) presented  
17 a high risk of contamination, since schools contained both treatment and control participants (e.g., Cook et al.  
18 2014; Smith 2004).

25 Selective outcome reporting presented the lowest risk of bias with 86% of included studies identified as low  
26 risk, judged by comparing protocols with final reports. If the final report produced data for the same variables  
27 originally proposed to be tested, we concluded there was a low risk of bias (e.g., Bradshaw et al. 2012; Lewis et  
28 al. 2013; Obsuth et al. 2016). It is important to mention that many of our included studies were unpublished  
29 PhD theses (i.e., 18 out of 37). As a result, pre-registration documents for these studies were often unavailable.  
30 We observed that most of the theses reported small or null effects (e.g., Arter 2005; Bragdon 2010; Burcham  
31 2002; Crowder 2001; Harding 2011; Peck 2006). Since null and small effects were reported, we assumed a low  
32 risk of bias regarding “selective outcome reporting”.

42 Arguably the seven methodologically strongest studies were Bradshaw et al. (2012), Cook et al. (2014), Hirsch  
43 et al. (2011), Johnson (1983), Lewis et al. (2013), Obsuth et al. (2016) and Wyman et al. (2010). All of them  
44 presented low risk in the randomisation process and most of them were clustered studies.

## 53 **Meta-analysis**

### 55 **Primary outcome: overall impact of school-based intervention**

57 The present analysis incorporates 38 effect sizes across 37 studies producing enough statistical information for  
58 meta-analysis (see Table 4). These studies represent a total sample of 37,895 students ( $Mage=12.5$ ,  $SD=2.85$ )

partaking in completed trials as treatment or control groups. On average, school exclusion was significantly reduced in the treatment group compared with the control group, post-treatment (i.e., six months on average). Under a random effects model, the standardised mean reduction was  $SMD=.30$ , 95%  $CI$  [.20, .41],  $p<.001$  meaning that those participating in school-based interventions were less likely to be excluded than those in the control group. Results exhibit significant heterogeneity ( $Q=301.3$ ;  $df=37$ ;  $p<.001$ ;  $I^2=87.7$ ;  $\tau^2=.078$ ) which was expected in this meta-analysis, bearing in mind that we include different school-based programmes, administering varying “doses” to participants in numerous locations and from different school grades.

As suggested by Lipsey and Wilson (2001:153) the U3 statistic for a  $SMD=.30$  is 62%. This would indicate that 62% of the treatment group is above the median of the control group. Stated differently, if we assume a 50/50 success rate for both groups, the treatment group has 62% success, versus 38% in the control group.

When we isolated the eleven studies measuring impact at follow-up (i.e., 12 or more months after finishing the intervention), the benefits of the interventions were less clear. In fact, the effect was reduced by half ( $SMD=.15$ , 95%  $CI$  [-.06, .35] and it was not statistically significant.

To increase precision in our results, we ran a meta-analysis with a subset of studies reporting both post-treatment and follow-up measures. Only seven studies reported short and long-term effect measures. At post-treatment, under a random effects model, the standardised mean reduction was  $SMD=.21$ , 95%  $CI$  [.11, .30]. However, when we ran the meta-analysis including only the subset of seven studies, the average time involved under “post treatment” was 12 months. This implies that although the overall effect is slightly lower than the general measure reported in Table 4, the impact lasts longer (i.e., 12 instead of 6 months on average). In the case of the effect at follow-up, the subset of studies produced an overall impact that was very small ( $SMD=.054$ ; 95%  $CI$  [-.04, .15] and non-significant. Heterogeneity was highly reduced ( $Q=7.80$ ;  $df=6$ ;  $p>.05$ ;  $I^2=23$ ;  $\tau^2=.004$ ) in the analysis of this subset of studies.

*TABLE 4 ABOUT HERE*

### Effectiveness by type of exclusion

Subgroup analysis using studies which reported different types of exclusion revealed that interventions were more effective at reducing in-school exclusion  $SMD=.35$ , 95%  $CI$  [.11, .58],  $p<.005$ ,  $n=6$  and expulsions  $SMD=.53$ , 95%  $CI$  [.07, .98],  $p<.05$ ,  $n=4$ . In the case of out-of-school exclusion, the impact of the intervention was close to zero and non- statistically significant  $SMD=.02$ , 95%  $CI$  [-.16, .19],  $p>.005$ ,  $n=9$ .

Finally, 27 studies presented data on suspension as a broad and general measure. These studies did not report operational definitions about the type of disciplinary measures involved in the disciplinary measure. With the aim of transparency, we report these results separately, although this measure could involve any of the previous outcomes reported above and, could therefore, be a subset of the overall effect size reported at the beginning of this section. The effect of school-based interventions was positive on General Suspension ( $SMD= .32$ , 95%  $CI$  [.21, .43],  $p<.001$ ), and like the overall effect size reported in Table 4 above, heterogeneity remained substantial ( $Q=171.4$ ;  $df=27$ ;  $p<.001$ ;  $I^2=84$ ;  $\tau^2=.056$ ).

### Overall effect on behaviour

For any study reporting data on school exclusion, we also coded secondary outcomes referring to internalising and externalising behaviours.

Statistical results on internalising behaviour were reported in only five trials and the data was in many cases insufficient for metacalculations. However, the narrative description suggests that school-based interventions had a small effect on the reduction of the above symptoms.

Regarding externalizing behaviour, 14 studies reported complete data for a composite measure of antisocial behaviour. The fourteen studies provided 15 independent effect sizes. Unusually, Wyman et al. (2010) reported a measure of behaviour control (e.g., children accepting imposed limits). We reversed the results as a proxy of antisocial behaviour and included this information in our calculations. The same procedure was followed with Feindler et al. (1984), who reported a measure of increase in self-control. Once again, we reversed the effects size as a proxy of antisocial behaviour.

The overall impact of school-based interventions on antisocial behaviour, under a random effects model, was not statistically different from zero  $SMD= - .005$ , 95%  $CI$  [-.097, .09],  $p >.05$  indicating an overall null effect of

these programmes in reducing antisocial behaviour.

It must be highlighted that some of the included studies in this overall measure reported negative effect sizes, meaning that in some specific cases the intervention made outcomes worse (e.g., Hawkins et al. 1988; Hostetler and Fisher 1997; Obsuth et al. 2016). For instance, Hawkins et al. (1988) reported the impact of an intervention focused on interactive teaching and co-operative learning targeting low achievers in mainstream schools. The treatment group showed a reduction in the number of exclusions but an increase in serious crime. Similar evidence was found by Hostetler and Fisher (1997) and Obsuth et al. (2016).

### **Moderator analysis**

Based on previous research, we coded a few moderator variables that could potentially explain heterogeneity in the overall effect size. Table 5 summarises the results of those analyses.

#### *TABLE 5 ABOUT HERE*

*Gender of the school population.* Eleven studies were implemented in schools where more than 60% of students were male, and 19 effects were tested in schools presenting a mixed population (i.e., neither gender exceeded 60 per cent). As shown in Table 5, interventions implemented in schools with predominantly male populations reported effects twice as larger as those targeting mixed-sex schools. However, differences between groups were not statistically significant, probably because of the small number of studies included.

*Age.* In terms of age, the best proxy available was school grade. To test the hypothesis that effect sizes vary by age, we ran subgroup analysis for 12 reports involving students from the six first years of education, 16 targeting years 7 to 9 and eight targeting students from years 10 to 12. The effect was notable larger in the last group, but the between effect difference was not statistically significant.

*Theoretical basis of interventions.* Information on the theoretical basis of interventions was not comprehensive. Studies in our meta-analysis tended to describe components of interventions more accurately than the theory or set of theories framing the approach. We therefore decided that the level where change was expected to occur would produce more consistent data on the theory underlying interventions. We then divided studies into those targeting a systemic change versus those targeting a change in students. The standardised mean difference for

the 10 evaluations targeting a systemic change (i.e., at the school level) provided a similar reduction in exclusion when compared with the 28 evaluations targeting a change at the pupil's level. Both independent effects were similar and statistically significant; however, the between-group comparison was non-significant, meaning that there is no evidence that the effect differs by level of targeted change (see Table 5).

*Quality of intervention: training and monitoring.* Previous research demonstrates that well-implemented programmes – those including training and monitoring – yield larger and more consistent effect sizes (e.g., Durlak et al. 2011; Gottfredson and Wilson 2003; Lösel and Beelmann 2006) than those lacking both components. Twenty-five studies (reporting 26 independent effect sizes) clearly stated the presence of training hours before intervention delivery. In the remaining 12 studies, the authors did not mention training. We ran sub-group analyses to test the speculative hypothesis that those reporting training would produce significantly improved outcomes. Both effects were close to  $SMD = .30$ , but the test of the difference between the two subgroups of studies yielded no evidence that the effect differs according to the presence or absence of prior training.

Fifteen studies reported monitoring the implementation of the programme during the trial. These yielded a result equal to  $SMD = .20$ , 95% CI [.05, .35],  $p < .05$ . In parallel the 23 studies that did not report monitoring produced an  $SMD = .37$ , 95% CI [.25, .50],  $p < .001$ . Both results were positive and statistically significant. The between-group comparison was non-significant, meaning that there is no convincing evidence that the effect differs by quality of intervention in these studies.

*Reasons for conducting the research.* Based on Singal et al. (2014) we compare the variation of effect between demonstration studies (i.e., studies testing the impact of an intervention under highly controlled optimal conditions) and routine evaluation studies (i.e., testing established programmes under circumstances that approach real-life conditions). As shown in Table 5, the 16 studies conducted for demonstration purposes reported an impact three times as larger as the 18 studies carried out for routine evaluation. The between-studies comparison was significant, meaning that the effect varied depending on the stage of research (i.e. larger effects under optimal conditions).

*Cluster versus individual level studies.* In the present review, we included primary studies involving individually randomised as well as cluster-randomised unit (e.g., schools or classrooms). To test the hypothesis

that the two subgroups would report different effect sizes, we run a moderator analysis. Thirteen effect sizes were calculated with data representing clusters of students, and twenty-five effects were based in data representing individuals. Results suggest that cluster studies reported a smaller effect than individual level reports. However, as observed in Table 5, differences between groups were not statistically significant.

*Evaluator role.* We coded data identifying the role of the evaluator. ‘Independent evaluators’ were those not taking part in the design or implementation of the evaluated programme. ‘Dependent evaluators’ were those also contributing to the design and/or the implementation of the programme. The 18 trials carried out by independent evaluators (i.e., those not taking part in the design or implementation of the evaluated programme) produced an effect size less than one-third as large as the 16 RCTs conducted by dependent evaluators (i.e., those who also developed and/or designed the intervention). The between-groups comparison was statistically significant, meaning that part of the variation in the overall effect size might be explained by the evaluator role.

*Type of the intervention.* As observed in Table 5, in this meta-analysis a small number of studies was included for each type of intervention, and for this reason, these results should be cautiously interpreted. Nevertheless, the standardised mean differences of only five types of programmes present positive (small to moderate effect sizes) and statistically significant results in favour of the reduction of school exclusion. Those programmes are: i) enhancement of academic skills; ii) mentoring/monitoring; iii) skills training for teachers; iv) counselling/mental health services; and v) other programmes. Since “other programmes” involve a mixture of different interventions, we believe they cannot be interpreted in the same way as the remaining four types. The comparison demonstrates that the between groups differences are statistically significant meaning that the variation in effect sizes might be explained to some extent by the type of intervention implemented.

### **Meta-regression**

To explore heterogeneity, we ran meta-regression using the moderators defined *a priori*. Model I included participants’ characteristics only, to get a sense of the net effect of these variables on the results. The gender of the school population, age (i.e., grade at school) and ethnicity (i.e., percentage of White people), as reported in Table 6, did not explain heterogeneity in the study results. In model II, we introduced the intervention characteristics (i.e., interventions targeting a change at the individual level versus those targeting a change at the

1 school level). Again, none of these variables were significant. This suggests that variability across effect sizes  
2 cannot be explained by our *a priori* moderators.  
3  
4

5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

*TABLE 6 ABOUT HERE*

Based on descriptive analysis of data, and also based on previous research findings (e.g., Beelmann and Lösel 2006; Farrington et al. 2016), we selected three post-hoc moderators, namely: i) reasons for conducting the research; ii) evaluator role; and iii) risk of bias. The results are described in Table 7.

*TABLE 7 ABOUT HERE*

The results were significant ( $p < .05$ ) only for the role of the evaluator. That coefficient is negative with the reference category “dependent”, meaning that the effect is lower when an independent team runs the research.

**Publication bias analysis**

As originally proposed, we used statistical procedures to quantify the potential publication bias that could affect our analysis.

Results of Duval and Tweedie's trim-and-fill analysis suggest that there were no differences in effect sizes attributable to bias. Under a fixed effect model, the point estimate for combined studies did not differ when comparing the original and the adjusted estimate (in both cases it was  $SMD = .39$ , 95% CI [.37, .42]). The same was true for the random effects model (in both cases they were  $SMD = .30$ , 95% CI [.20, .41]). Based on the parameter of Duval and Tweedie's trim-and-fill, it seems that no studies are missing. Rosenthal's fail-safe N test indicated that it would be necessary to allocate and include 2,347 missing studies to nullify the observed effect, which seems highly unlikely.

**Sensitivity analysis**

In the present meta-analysis, the effect of the treatment was calculated as the difference between post-treatment and baseline for treatment and control groups. We corrected the value of the variance by assuming a value of correlation equals to .75. To test the robustness of this assumption, we ran sensitivity analysis with a value of .50. Table 8, panel A shows that overall results remain stable when the correlation is smaller.

Another decision was related to the presence of outliers. We found one study presenting an effect size more than three standard deviations from the mean effect size, which was defined as an outlier (Collier 2002). We tested the impact of the outlier and the impact of windsorisation. The size of the effects, their direction and significance were not altered. See Table 8 below, panels B and C.

#### TABLE 8 ABOUT HERE

Finally, as stated in the protocol, we ran a sensitivity analysis to test differences between published and unpublished reports. The 20 independent effect sizes reported in 19 peer-reviewed journals yield a  $SMD=.31$ , 95%  $CI$  [.17, .45], whereas the 18 effect sizes extracted from unpublished reports reported a  $SMD=.29$ , 95%  $CI$  [.11, .47]. The between effect difference was not statistically significant.

## DISCUSSION

The results from this meta-analysis provide some evidence of short term effectiveness of school-based interventions in reducing school exclusion. Although our empirical results must be cautiously considered, they suggest that school administrators and policymakers do have alternatives to exclusion when dealing with disciplinary problems.

The nine types of interventions from the included studies were compared to test the hypothesis that some may be more effective than others. Enhancement of academic skills, mentoring/monitoring, skills training for teachers and counselling/mental health services reported the largest and most significant effect sizes. Based on the number of studies included in each sub-type, it is our judgement that skills training for teachers and mentoring/monitoring represent the strongest and more reliable results. These results are in line with previous research that emphasizes the importance of teachers' skills and mentoring programmes in promoting prosocial behaviours and values (Freiberg and Lapointe 2006; Oliver et al. 2011; Tolan et al. 2008). Tolan et al. (2008), for instance, found that mentoring was effective in reducing delinquency, aggression and drug use, and improving academic achievements (albeit with small effects). Furthermore, mentoring was more effective when



(a) participants had greater pre-existing behavioural problems or had been exposed to significant levels of environmental risk, (b) they were male, (c) the educational or occupational backgrounds of the mentors fitted the goals of the program, (d) mentors and youths were successfully paired, with similar interests, and (e) programmes were structured to support mentors. In addition, the meta-analysis conducted by Eby et al. (2008) also found that mentoring programmes had a small and significant positive effect on academic performance.

As far as teachers' skills for managing students' behaviour is concerned, previous evidence suggests that such programmes can improve teachers' general instructional and behavioural management skills in planning, implementing and maintaining effective classroom practices. In line with our results, a meta-analysis carried out by Oliver et al. (2011) concluded that classroom management practices had a significant, positive effect on reducing problem behaviour. Students taking part in the intervention displayed less disruptive, inappropriate and aggressive behaviour in the classroom compared to those in control classrooms.

The positive relationship between teachers and students has been found to be a factor in promoting more prosocial and less aggressive behaviours later in life. A recent study by Obsuth et al. (2016:16), found that: "teachers who reported having a more positive relationship with a student at age ten observed significantly fewer aggressive and defiant behaviours and more prosocial behaviours in the same student concurrently and one year later, at age 11. This was also associated with more prosocial behaviours two years later, at age 12 and also with less aggressive behaviour at age 13. Similarly, students who perceived a more positive relationship with their teacher at age 11 reported fewer aggressive behaviours and more prosocial behaviours concurrently and fewer aggressive behaviours two and four years later, at ages 13 and 15."

These findings are in line with previous meta-analyses consistently stating that the quality of the student-teacher relationship is a strong predictor of classroom behaviour, students' motivation, engagement to school and achievements at all ages (Allen et al. 2007; Cornelius-white 2007; Roorda et al. 2011). Even if most of the previous evidence is focused on the first years of schooling, recent research suggests that the effect of student-teacher positive attachment remains relevant during adolescence. As stated by Okonofua et al. (2016: 5221) "Relationships of trust and respect may be especially important in adolescence. In this period before cognitive-control regions in the brain have fully matured, external resources like trusted teachers may be essential to guide children's growth". On the contrary, punitive interactions, based on exclusion and labelling some students as

problematic, may produce a never-ending cycle of punishment and misbehaviour (Matsueda 2014; Okonofua et al. 2016; Sherman 2010; Tyler and Huo 2002; Way 2011). All these results make us believe that investing in teachers' skills and positive relationships between students and teachers is worthwhile, with schools becoming target locations for preventing crime and promoting positive psychosocial development.

As originally planned, we assessed the impact of school-based interventions on pupil behaviour, a secondary outcome domain. It was hypothesised that a reduction in exclusion would be linked with variations in students' behaviours. Internalising behaviours were only reported in a small number of studies and it was impossible to produce a pooled effect size. For externalising behaviours (i.e., delinquency, violence, bullying, antisocial behaviour and a minority of studies reporting oppositional behaviour) the results showed an impact close to zero that was non-significant. It follows that interventions aimed at reducing exclusion do not necessarily reduce antisocial behaviour. This result is in line with evidence suggesting that changes in school policies, rather than changes in behaviour, produce reductions in disciplinary exclusion (Noltemeyer and Fenning 2013; Skiba et al. 2015). One could hypothesise that, if ethnicity is a stronger predictor of school exclusion than other demographic and behavioural characteristics, antisocial behaviour would not necessarily explain rates of exclusion. What is more, it would not be the main cause of this punishment being used. However, based on our data, these ideas cannot yet be regarded as conclusive and more research is needed about the causes of exclusion.

Pre-determined moderators (i.e., participants' characteristics, the theoretical basis of the interventions, and the quality of the interventions) were not effective in explaining the heterogeneity in our results. Even when the impact of interventions was three times larger in schools with predominantly male populations than mixed-sex schools, the between-group comparison was non-significant, probably because of small number of studies. Given the body of literature pointing out the higher base-rates of problem behaviour in males, we believe that future studies *could* make these differences significant, but based on our findings, we could not confirm the hypothesis that the effect differs by school gender. The large effect size is more noteworthy than the lack of statistical significance, which is caused by small N.

Among three *post-hoc* moderators, the role of the evaluator was found to be significant: independent evaluator teams reported lower effect sizes than research teams who were also involved in the design and/or delivery of the intervention. This is not surprising and there is extensive research documenting this phenomenon as well as

the advantages of running independent evaluations (e.g., Eisner 2009; Eisner and Humphreys, 2012; Lösel and Beelmann, 2006b; Petrosino and Soydan, 2005; Wilson et al. 2003). However, in the interest of balance, one could speculate that this could reflect the fact that the dependant research teams implemented and evaluated the intervention better.

## Limitations

Our study includes limitations that warrant readers' attention. First, it is important to acknowledge that even though we focused on randomised controlled trials, the included studies present limited information for judging bias. More than 50% of the reports failed to provide enough data to judge the adequacy of randomisation, that is, sequence generation and allocation concealment. As recently stated by Roberts and Ker (2016), missing data on those details can drive analysts to identify "false positive" RCTs. We cannot claim that this was the case among our studies, but clearly the absence of detail imposes limitations on the assessment of quality in regard to potential threats to internal validity. A similar lack of data affected the evaluation of blind outcome measures. Even if most of the exclusion measures in our report are based on official records, we cannot overlook the fact that teachers or school staff are in charge of imposing sanctions and could potentially be aware of a student's participation in the experimental evaluation.

Secondly, 35% of included studies were based on samples with less than 100 participants. Small samples impose clear limitations on the ability to detect the effects of interventions. Future research would benefit from prospective power analysis to ensure adequately powered studies (Ellis 2010). Future meta-analyses should also give more attention to this issue and, as suggested by Farrington et al. (2016), it seems advantageous to set a minimum sample size for inclusion in reviews.

Third, as reported in the moderator analysis, the independence of the evaluator explained the heterogeneity of effect sizes. Even if independent teams reported on a good number of our studies, close to 50% of evaluations did not use independent evaluators. This fact alone could increase bias.

However, the present study also has several strengths. First, this review and meta-analysis is the first attempt to collect and statistically summarise interventions intended to reduce school exclusion. As such, we believe that our report offers an overview of the amount, characteristics, limitations, and quality of the available evidence, as

well as a measure of the size of the effect achieved by different types of intervention. Secondly, we have endeavoured to use an exhaustive coding process to provide careful descriptions of the studies as well as a meticulous analysis of the statistical data available. Overall, this review represents the best available evidence on the effectiveness of interventions aimed at reducing school exclusion.

## Future research

The results of this first review on the topic identify some implications for future research, laid out below:

*Addressing the ‘ethnicity gap’ in exclusions.* Most of the literature that was reviewed indicated that racial or ethnic identity plays a central role in predicting school exclusion, even after controlling for demographic and behavioural variables. More research needs to be developed for testing the mechanisms that produce different treatment for some racial minorities<sup>2</sup>. It is necessary to understand the individual and social processes that lead to the overrepresentation of minorities in school exclusion. As suggested by some scholars, the divergence between students and teachers’ cultural expectations could potentially contribute to misinterpretation of each other’s behaviour, fears and conflicts (Gregory et al. 2010). Understanding the mechanisms that make race a predictor of exclusion could have implications for future policy and practice. As stated by Irvin (2002) in previous research, greater diversity among staff could be helpful in promoting understanding of cultural differences and reducing bias.

*Causal effects of exclusion.* A review of previous research suggests that the causal effect of punishment on students’ behaviour is still a long way from being fully understood (Sulger et al. 2013). If school exclusion is simply a marker of an underlying antisocial syndrome, it could be beneficial to invest in prevention programmes targeting, for example, deviant behaviours (e.g., violence, drug use, crime, abuse, and neglect) and personality features (e.g., aggressiveness, lack of empathy, lack of remorse) associated with the antisocial syndrome (Farrington 2003). However, if school exclusion is proved to be the *cause* of detrimental outcomes later in life, it will be worth investing in more programmes focusing specifically on the reduction of school exclusion.

---

<sup>2</sup> For further details see van den Bergh et al. 2010 on the concept of implicit bias.

1 *Mediating mechanisms that explain reductions/increases in exclusion rates.* Future evaluations of school-based  
2 interventions aimed at reducing exclusion need to explore the presumed causal mechanisms that lead to that  
3 reduction. More theoretically informed trials that examine causal mechanisms are important in order to design  
4 better interventions. In fact, this information would be crucial in planning future prevention programmes since  
5 causal mechanisms can shed light on what works for whom and under what conditions.  
6  
7  
8  
9

10  
11 *Attempt blind assessment in school-based RCTs.* The characteristics of schools make it hard to blind all those  
12 involved in trials. Teachers, students and school counsellors are likely to be involved in the experiment, making  
13 it difficult to control the social desirability effect of those participating in the study (Hutchison and Styles 2010).  
14 This challenge needs to be addressed by future studies by at least blinding those who collect outcome data. It  
15 also seems necessary, at the level of meta-analytical studies, that the tools used to measure quality bias in school  
16 contexts be adapted. In our experience, some of the available “risk of bias tools” seemed more suited for  
17 medical trials than for school-based experiments and we believe there is room to make these more suitable for  
18 social research.  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29

30 *Risk of quality bias.* By the fact that meta-analysis combines results from different primary research reports,  
31 quality bias involved in primary research can jeopardise the validity of the “meta” results. Differences in the  
32 quality of primary studies can result in false positive conclusions when less rigorous studies are biased toward  
33 overestimating an intervention’s effect. These can also be false negative conclusions in those cases where less  
34 rigorous studies are inclined towards underestimating an effect Higgins and Green (2011:189).  
35  
36  
37  
38  
39  
40  
41

42 Future research would benefit from following CONSORT standards. Registration of trials, even if they are PhD  
43 theses would represent a huge benefit not only for the scientific community, but also for those interested in  
44 evidence-based decision making.  
45  
46  
47  
48  
49

50 *Cross-cultural research.* Based on our findings, the evidence so far has largely come from the United States.  
51 More research needs to be done in other countries where school exclusion is an issue. We know that evidence  
52 suggesting effective approaches in some countries/cultures will not necessarily have the same effectiveness  
53 when translated to different populations. Those making decisions about how to reduce exclusion in their own  
54 country need to have access to detailed information addressing their needs.  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

*Innovative strategies*. More research needs to be conducted on innovative strategies, for example those involving empathy-based philosophies. This was the basis for the intervention tested by Okonofua et al. (2016: 5521) included in this review. The intervention was focused on encouraging teachers to adopt an empathetic attitude towards discipline; it is low-cost, and demonstrates long-term effects (rates of exclusion were reduced by 50% during one academic year). As stated by the authors, “teachers’ mind sets about discipline directly affect the quality of teacher-student relationships and student suspensions and, moreover, can be changed through scalable intervention” .

## CONCLUSIONS

The empirical evidence produced by this review suggests that school-based programmes can reduce the use of exclusion. Even if the effects are not sustained in the long-term, the data shows that in the short-term schools can opt for effective alternative approaches to managing discipline, rather than exclusion. This review, aimed at testing the effectiveness of school-based programmes, offers a broad overview not only of the effectiveness of different interventions, but also uncovers findings that can guide public policy and future research.

It is important to emphasize that the results are most applicable to the United States, where the majority of the assessed evidence was collected. As expressed in the previous paragraphs, some flaws affect the primary evaluations included in this review. Nevertheless, we believe that our conclusions represent the best possible synthesis of the currently available evidence. We hope that results encourage researchers to produce better quality evidence rather than abandoning their efforts to find strategies to replace exclusionary punishments.

## REFERENCES

*References marked with an asterisk indicate studies included in the meta-analysis.*

- Achenbach, T. M. (1978). The child behaviour profile: I. Boys aged 6–11. *Journal of Consulting and Clinical Psychology*, 46(3), 478–488.
- Achenbach, T. M., & Edelbrock, C. S. (1979). The Child behaviour Profile: II Boys Aged 12-16 and Girls 6-11 and 12-16. *Journal of Consulting and Clinical Psychology*, 47(2), 223–233. <http://doi.org/10.1037/0022-006X.52.2.207>
- Achilles, G. M., McLaughlin, M. J., & Croninger, R. G. (2007). Sociocultural correlates of disciplinary exclusion among students with emotional, behavioural, and learning disabilities in the SEELS national dataset. *Journal of Emotional and behavioural Disorders*, 15(1), 33–45. <http://doi.org/10.1177/10634266070150010401>
- Ahn, S., Myers, N. D., & Jin, Y. (2012). Use of the estimated intraclass correlation for correcting differences in effect size by level. *behaviour Research Methods*, 44(2), 490–502. <http://doi.org/10.3758/s13428-011-0153-1>

- 1 \*Allen, J. P., Philliber, S., Herrling, S., & Kuperminc, G. P. (1997). Preventing teen pregnancy and academic  
2 failure: Experimental evaluation of a developmentally based approach. *Child Development*, 64(4), 729–  
3 742. <http://doi.org/10.2307/1132122>
- 4 Allen, J., Philliber, S., & Herrling. (1997). Preventing teen pregnancy and academic failure: Experimental  
5 evaluation of a developmentally based approach. *Child Development*, 64(4), 729–742.
- 6 Allen, M., Witt, P. L., Wheelless, L. R., Allen, M., Witt, P. L., & Wheelless, L. R. (2007). The role of teacher  
7 immediacy as a motivational factor in student learning: Using meta-analysis to test a causal model.  
8 *Communication Education*, 55(1), 21–31. <http://doi.org/10.1080/03634520500343368>
- 9 American Psychological Association Zero Tolerance Task Force [APA ZTTF]. (2008). *Are zero tolerance*  
10 *policies effective in the schools? An evidentiary review and recommendations*. *The American Psychologist*  
11 (Vol. 63). Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/19086747>
- 12 Arcia, E. (2006). Achievement and enrollment status of suspended students: Outcomes in a large, multicultural  
13 school district. *Education and Urban Society*, 38(3), 359–369. <http://doi.org/10.1177/0013124506286947>
- 14 \*Arter, P. (2005). *The Effects of the Positive Alternative Learning Supports Program on the Attendance,*  
15 *Academic Performance, and behaviour of Students with Chronic behaviour Problems*. The Johns Hopkins  
16 University, Maryland. Retrieved from <http://search.proquest.com/docview/304991742?accountid=9851>;  
17 [http://tf5lu9ym5n.search.serialssolutions.com/?ctx\\_ver=Z39.88-2004&ctx\\_enc=info:ofi/enc:UTF-](http://tf5lu9ym5n.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rft_id=info:sid/ProQuest+Dissertations+&+Theses+A&I&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation)  
18 [8&rft\\_id=info:sid/ProQuest+Dissertations+&+Theses+A&I&rft\\_val\\_fmt=info:ofi/fmt:kev:mtx:dissertation](http://tf5lu9ym5n.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rft_id=info:sid/ProQuest+Dissertations+&+Theses+A&I&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation)
- 19 \*Barnes, V. A., Bauza, L. B., & Treiber, F. A. (2003). Impact of stress reduction on negative school behaviour  
20 in adolescents. *Health and Quality of Life Outcomes*, 1(10), 1–7. <http://doi.org/10.1186/1477-7525-1-10>
- 21 Beelmann, A., & Lösel, F. (2006). Child Social Skills Training in Developmental Crime Prevention : Effects on  
22 Antisocial behaviour and Social Competence. *Psicothema*, 18, 603–610.
- 23 \*Berlanga, D. T. (2004). *The efficacy of an intensive guidance intervention program to increase academic*  
24 *success of at-risk middle school students*. Texas A&M University - Corpus Christi. Retrieved from  
25 <http://search.proquest.com/docview/305051136?accountid=9851>;  
26 [http://tf5lu9ym5n.search.serialssolutions.com/?ctx\\_ver=Z39.88-2004&ctx\\_enc=info:ofi/enc:UTF-](http://tf5lu9ym5n.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rft_id=info:sid/ProQuest+Dissertations+&+Theses+A&I&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation)  
27 [8&rft\\_id=info:sid/ProQuest+Dissertations+&+Theses+A&I&rft\\_val\\_fmt=info:ofi/fmt:kev:mtx:dissertation](http://tf5lu9ym5n.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rft_id=info:sid/ProQuest+Dissertations+&+Theses+A&I&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation)
- 28 Borenstein, M., Hedges, L., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*.  
29 Chichester, UK: John Wiley & Sons, Ltd.
- 30 Bowman-Perrott, L., Benz, M. R., Hsu, H.-Y., Kwok, O.-M., Eisterhold, L. A., & Zhang, D. (2013). Patterns  
31 and predictors of disciplinary exclusion over time: An analysis of the SEELS national data set. *Journal of*  
32 *Emotional and behavioural Disorders*, 21(2), 83–96. <http://doi.org/10.1177/1063426611407501>
- 33 \*Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2012). Effects of school-wide positive behavioural  
34 interventions and supports on child behaviour problems. *Pediatrics*, 130(5), e1136-45.  
35 <http://doi.org/10.1542/peds.2012-0243>
- 36 \*Bragdon, D. (2010). *The effects of a career awareness project on the academic performance of at-risk eighth*  
37 *grade students*. The Johns Hopkins University. Retrieved from  
38 <http://search.proquest.com/docview/1507474004?accountid=9851>;  
39 [http://tf5lu9ym5n.search.serialssolutions.com/?ctx\\_ver=Z39.88-2004&ctx\\_enc=info:ofi/enc:UTF-](http://tf5lu9ym5n.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rft_id=info:sid/ProQuest+Dissertations+&+Theses+A&I&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation)  
40 [8&rft\\_id=info:sid/ProQuest+Dissertations+&+Theses+A&I&rft\\_val\\_fmt=info:ofi/fmt:kev:mtx:dissertation](http://tf5lu9ym5n.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rft_id=info:sid/ProQuest+Dissertations+&+Theses+A&I&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation)
- 41 \*Brett, P. F. (1993). *An examination of the impact of a mentoring program on inner city, seventh-grade, African*  
42 *American students' attitudes toward learning and other selected variables*. The George Washington  
43 University. Retrieved from <http://search.proquest.com/docview/288249882?accountid=9851>;  
44 [http://tf5lu9ym5n.search.serialssolutions.com/?ctx\\_ver=Z39.88-2004&ctx\\_enc=info:ofi/enc:UTF-](http://tf5lu9ym5n.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rft_id=info:sid/ProQuest+Dissertations+&+Theses+A&I&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation)  
45 [8&rft\\_id=info:sid/ProQuest+Dissertations+&+Theses+A&I&rft\\_val\\_fmt=info:ofi/fmt:kev:mtx:dissertation](http://tf5lu9ym5n.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rft_id=info:sid/ProQuest+Dissertations+&+Theses+A&I&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation)
- 46 Brookes, M., Goodall, E., & Heady, L. (2007). *Misspent youth: The cost of truancy and exclusion*. *A guide for*  
47 *donors and funders*. New Philanthropy Capital. London.
- 48 Brown, T. M. (2007). Lost and turned out: Academic, social, and emotional experiences of students excluded  
49 from school. *Urban Education*, 42(5), 432–455. <http://doi.org/10.1177/0042085907304947>
- 50 \*Burcham, B. . (2002). *Impact of school-based social problem-solving training on middle school students with*  
51 *disruptive behaviour*. University of Kentucky. Retrieved from  
52 <http://search.proquest.com/docview/305510149?accountid=9851>;  
53 [http://tf5lu9ym5n.search.serialssolutions.com/?ctx\\_ver=Z39.88-2004&ctx\\_enc=info:ofi/enc:UTF-](http://tf5lu9ym5n.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rft_id=info:sid/ProQuest+Dissertations+&+Theses+A&I&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation)  
54 [8&rft\\_id=info:sid/ProQuest+Dissertations+&+Theses+A&I&rft\\_val\\_fmt=info:ofi/fmt:kev:mtx:dissertation](http://tf5lu9ym5n.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rft_id=info:sid/ProQuest+Dissertations+&+Theses+A&I&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation)
- 55 Burrell, N. A., Zirbel, C. S., & Allen, M. (2003). Evaluating peer mediation outcomes in educational settings: A  
56 meta-analytic review. *Conflict Resolution Quarterly*, 21(1), 7–26. <http://doi.org/10.1002/crq.46>
- 57 Chin, J. K., Dowdy, E., Jimerson, S. R., & Rime, W. J. (2012). Alternatives to Suspensions: Rationale and  
58 Recommendations. *Journal of School Violence*, 11(2), 156–173.  
59 <http://doi.org/10.1080/15388220.2012.652912>
- 60 Cochrane Effective Practice and Organisation of Care (EPOC). (2007). Suggested risk of bias criteria for EPOC

reviews. *EPOC Resources for Review Authors*.

- Cohen, J. (1960). A Coefficient for Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), 37–46. <http://doi.org/10.1177/001316446002000104>
- \*Collier, C. E. (2002). *The effect of pro-social skills training on the problem behaviour of selected African American children in the District of Columbia public schools: Recommendations for change*. The George Washington University, US. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2002-95010-383&site=ehost-live&scope=site>
- \*Cook, P. J., Dodge, K., Farkas, G., Fryer Jr, R. G., Guryan, J., Ludwig, J., ... Steinberg, L. (2014). The (Surprising) Efficacy of Academic and behavioural Intervention with Disadvantaged Youth from a Randomized Experiment in Chicago. Institute for Policy Research Northwestern University. Evanston, IL: Institute for Policy Research Northwestern University. <http://doi.org/10.3386/w19862>
- Cornelius-white, J. (2007). Learner-centered teacher-student relationships are effective: A meta-analysis. *Review of Educational Research*, 77(1), 113–143.
- \*Cornell, D. G., Allen, K., & Fan, X. (2012). A Randomized Controlled Study of the Virginia Student Threat Assessment Guidelines in Kindergarten Through Grade 12. *School Psychology Review*, 41(1), 100–115.
- Cornell, D. G., Gregory, A., & Fan, X. (2011). Reductions in long-term suspensions following adoption of the Virginia Student Threat Assessment guidelines. *NASSP Bulletin*, 95(3), 175–194. <http://doi.org/10.1177/0192636511415255>
- Costenbader, V., & Markson, S. (1998). School Suspension: A Study with Secondary School Students. *Journal of School Psychology*, 36(1), 59–82. Retrieved from <http://libsta28.lib.cam.ac.uk:2070/science/article/pii/S0022440597000502?np=y>
- \*Crowder, D. . (2001). *The effect of the gang resistance education and training program on character education and the development of leadership skills in children at an inner-city middle school*. Tennessee State University. Retrieved from [http://search.proquest.com/docview/287860152?accountid=9851;http://tf5lu9ym5n.search.serialssolutions.com/?ctx\\_ver=Z39.88-2004&ctx\\_enc=info:ofi/enc:UTF-8&rft\\_id=info:sid/ProQuest+Dissertations++Theses+A&I&rft\\_val\\_fmt=info:ofi/fmt:kev:mtx:dissertation](http://search.proquest.com/docview/287860152?accountid=9851;http://tf5lu9ym5n.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rft_id=info:sid/ProQuest+Dissertations++Theses+A&I&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation)
- Department for Education [DfE]. (2012). *A profile of pupil exclusions in England*. London.
- Department for Education [DfE]. (2013). Statistical first release: Permanent and fixed period exclusion from schools and exclusion appeals in England, 2011/12. *Department of Education SFR 29/2013*, 12. Retrieved from [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/224893/SFR29-2013.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/224893/SFR29-2013.pdf)
- Department for Education [DfE]. (2017). Permanent and fixed period exclusions in England: 2015 to 2016. *National Statistics, SFR(35/2017)*, 1–9.
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions. *Child Development*, 82(1), 405–432. <http://doi.org/10.1111/j.1467-8624.2010.01564.x>
- \*Dynarski, M., Moore, M., Mullens, J., Gleason, P., James-Burdumy, S., Rosenberg, L., ... Levy, D. (2003). *When Schools Stay Open Late: The National Evaluation of the 21st-Century Community Learning Centers Program. First Year Findings*. Mathematica Policy Research, Inc. Decision Information Resources, Inc. Retrieved from U.S. Department of Education
- Eby, L. T., Allen, T. D., Evans, S. C., Ng, T., & Dubois, D. L. (2008). Does Mentoring Matter? A Multidisciplinary Meta-Analysis Comparing Mentored and Non-Mentored Individuals. *Journal of Vocational behaviour*, 72(2), 254–267. <http://doi.org/10.1016/j.jvb.2007.04.005>
- \*Edmunds, J. A., Bernstein, L., Unlu, F., Glennie, E., Willse, J., Smith, A., & Arshavsky, N. (2012). Expanding the Start of the College Pipeline: Ninth-Grade Findings From an Experimental Study of the Impact of the Early College High School Model. *Journal of Research on Educational Effectiveness*, 5(2), 136–159. <http://doi.org/10.1080/19345747.2012.656182>
- Eisner, M. (2009). No effects in independent prevention trials: Can we reject the cynical view? *Journal of Experimental Criminology*, 5, 163–183. <http://doi.org/10.1007/s11292-009-9071-y>
- Eisner, M., & Humphreys, D. K. (2012). Measuring Conflict of Interest in Prevention and Intervention Research. A Feasibility Study. In T. Bliesener, A. Beelmann, & M. Stemmler (Eds.), *Antisocial behaviour and crime: Contributions of developmental and evaluation research to prevention and intervention*. Göttingen, London: Hogrefe Publishing.
- Ellis, P. D. (2010). *The Essential Guide to Effect Sizes. Statistical Power, Meta-Analysis, and the Interpretation of Research Results*. Cambridge, UK: Cambridge University Press.
- Fabelo, T., Thompson, M. D., Plotkin, M., Carmichael, D., Marchbanks, M. P., & Booth, E. A. (2011). *Breaking Schools' Rules: A Statewide Study of How School Discipline Relates to Students' Success and Juvenile Justice Involvement*.
- \*Farrell, A. D., Meyer, A. L., & White, K. S. (2001). Evaluation of Responding in Peaceful and Positive ways



- (RIPP): A school-based prevention program for reducing violence among urban adolescents. *Journal of Clinical Child Psychology*, 30(4), 451–463.
- Farrington, D. P. (1989). Early predictors of adolescent aggression and adult violence. *Violence and Victims*, 4(4), 79–100.
- Farrington, D. P. (2003). Advancing knowledge about the early prevention of adult antisocial behaviour. In D. P. Farrington & J. W. Coid (Eds.), *Early Prevention of Adult Antisocial behaviour* (pp. 1–20). Cambridge, UK: Cambridge University Press.
- Farrington, D. P., Ttofi, M. M., & Lösel, F. (2016). Developmental and social prevention. In D. Weisburd, D. P. Farrington, & C. Gill (Eds.), *What Works in Crime Prevention and Rehabilitation. Lessons from Systematic reviews* (pp. 15–75). New York: Springer Series on Evidence-Based Crime Policy.
- Farrington, D., & Ttofi, M. M. (2009). School-based Programs to Reduce Bullying and Victimization. *Campbell Systematic Reviews*, 6. <http://doi.org/10.4073/csr.2009.6>
- \*Feindler, E., Marriott, S., & Iwata, M. (1984). Group Anger Control Training for junior high school delinquents. *Cognitive Therapy and Research*, 8(3), 299–311. <http://doi.org/10.1007/BF01173000>
- Freiberg, H. J., & Lapointe, J. M. (2006). Research-based programs for preventing and solving discipline problems. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management research, practice, and contemporary issues* (pp. 735–786). London: Routledge. <http://doi.org/10.4324/9780203874783.ch28>
- Gottfredson, D. C., & Wilson, D. B. (2003). Characteristics of Effective School-based Substance Abuse Prevention. *Prevention Science*, 4(1), 27–38. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12611417>
- Gregory, A., Skiba, R. J., & Noguera, P. A. (2010). The achievement gap and the discipline gap: Two sides of the same coin? *Educational Researcher*, 39(1), 59–68. <http://doi.org/10.3102/0013189X09357621>
- \*Harding, H. (2011). *An evaluation of a targeted group intervention delivered to year 8 pupils and broadly based on cognitive-behavioural approaches*. University of Nottingham. Retrieved from <http://eprints.nottingham.ac.uk/11843/>
- \*Hawkins, J. D., Doueck, H. J., & Lishner, D. M. (1988). Changing Teaching Practices in Mainstream Classrooms to Improve Bonding and behaviour of Low Achievers. *American Educational Research Journal*, 25(1), 31–50. <http://doi.org/http://dx.doi.org/10.3102/00028312025001031>
- Hedges, L. V. (2007). Effect Sizes in Cluster-Randomized Designs. *Journal of Educational and behavioural Statistics*, 32(4), 341–370. <http://doi.org/10.3102/1076998606298043>
- Hemphill, S. A., & Hargreaves, J. (2010). School suspensions. A resource for teachers and school administrators. *Centre for Adolescent Health, Murdoch Childrens Research Institute, Melbourne*.
- Hemphill, Sheryl, A., Toumbourou, John, W., Smith, & Rachel. (2010). Are rates of school suspension higher in socially disadvantaged neighbourhoods? An Australian study. *Health Promotion Journal of Australia*, 21(1), 12–18. Retrieved from [http://apps.isiknowledge.com/full\\_record.do?product=UA&search\\_mode=GeneralSearch&qid=3&SID=Y1L6KdejMLbiGol1p&page=1&doc=7](http://apps.isiknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=3&SID=Y1L6KdejMLbiGol1p&page=1&doc=7)
- Higgins, J. P. T., & Green, S. (2011). *Cochrane Handbook for Systematic Reviews of interventions. Version 5.1.0*. Retrieved from [www.cochrane-handbook.org](http://www.cochrane-handbook.org)
- \*Hirsch, B. J., Hedges, L. V., Stawicki, J., & Mekinda, M. A. (2011). *After-School Programs for High School Students. An Evaluation of After School Matters*.
- \*Hostetler, M., & Fisher, K. (1997). Project CARE substance abuse prevention program for high-risk youth: A longitudinal evaluation of program effectiveness. *Journal of Community Psychology*, 25(5), 397–419. [http://doi.org/10.1002/\(sici\)1520-6629\(199709\)25:5<397::aid-jcop3>3.0.co;2-q](http://doi.org/10.1002/(sici)1520-6629(199709)25:5<397::aid-jcop3>3.0.co;2-q)
- Hostetler, M. L., & Vondracek, F. (1995). *Project C.A.R.E. substance abuse prevention program: An evaluation of program effectiveness*. The Pennsylvania State University, Ann Arbor. Retrieved from <http://search.proquest.com/docview/304210183?accountid=9851>
- Hutchison, D., & Styles, B. (2010). *A guide to running randomised controlled trials for educational researchers*. NFER Slough, UK.
- \*Ialongo, N., Poduska, J., Werthamer, L., & Kellam, S. (2001). The distal impact of two first-grade preventive interventions on conduct problems and disorder in early adolescence. *Journal of Emotional and behavioural Disorders*, 9(3), 146–160. <http://doi.org/10.1177/106342660100900301>
- Irvin, J. (2002). *In search of wholeness. African American teachers and their culturally specific classroom practices*. New York: Palgrave Macmillan. <http://doi.org/10.1057/9780230107182>
- \*Johnson, N. H. S. (1983). *The Effects Of Affective Education On The Discipline Of Selected Seventh And Eighth Grade Students In The District Of Columbia Public Schools*. The George Washington University, US. Retrieved from [http://search.proquest.com/docview/303265600?accountid=9851;http://tf5lu9ym5n.search.serialssolutions.com/?ctx\\_ver=Z39.88-2004&ctx\\_enc=info:ofi/enc:UTF-8&rft\\_id=info:sid/ProQuest+Dissertations+%26+Theses+A%261&rft\\_val\\_fmt=info:ofi/fmt:kev:mtx:disse](http://search.proquest.com/docview/303265600?accountid=9851;http://tf5lu9ym5n.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rft_id=info:sid/ProQuest+Dissertations+%26+Theses+A%261&rft_val_fmt=info:ofi/fmt:kev:mtx:disse)

rtatio

- \*Lewis, K. M., Schure, M. B., Bavarian, N., DuBois, D. L., Day, J., Ji, P., ... Flay, B. R. (2013). Problem behaviour and Urban, Low-Income Youth: A Randomized Controlled Trial of Positive Action in Chicago. *American Journal of Preventive Medicine*, 44(6), 622–30. <http://doi.org/10.1016/j.amepre.2013.01.030>
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical Meta-Analysis*. Thousand Oaks, California: Sage Publications.
- Liu, J. C. (2013). The suspension spike: Changing the discipline culture in NYC's middle schools. *New York City Comptroller's Office*, 1–65.
- Lösel, F., & Beelmann, A. (2006). Child Social Skills Training. In B. C. Welsh & D. P. Farrington (Eds.), *Preventing Crime. What works for Children, Offenders, Victims and Places*. (pp. 33–54). Dordrecht: Springer.
- \*Mack, M. A. P. (2001). *The effects of a behavioural intervention on discipline referrals and school suspension*. Auburn University. Retrieved from <http://search.proquest.com/docview/275621879?accountid=9851>
- Massey, A. (2011). Best Behaviour: School discipline, intervention and exclusion. *Policy Exchange*, 75. Retrieved from <http://www.policyexchange.org.uk/publications/category/item/best-behaviour-school-discipline-intervention-and-exclusion>
- Matsueda, R. L. (2014). The Natural History of Labeling Theory. In D. P. Farrington & J. Murray (Eds.), *Labelling Theory. Empirical Tests. Advances in criminological theory. Volume 18* (pp. 11–44). New Brunswick, New Jersey.: Transaction Publishers.
- Michail, S. (2011). Understanding school responses to students' challenging behaviour: A review of literature. *Improving Schools*, 14(2), 156–171. <http://doi.org/10.1177/1365480211407764>
- Ministry of Education Ontario Canada. (2014). Safe schools: Suspension and expulsion facts, 2013-2014. Retrieved November 3, 2016, from <http://www.edu.gov.on.ca/eng/safeschools/facts1314.html>
- Mytton, J. A., DiGiuseppi, C., Gough, D., Taylor, R. S., & Logan, S. (2009). School-based secondary prevention programmes for preventing violence (Review). *The Cochrane Collaboration*, (4).
- Nichols, J. D. (2004). An exploration of discipline and suspension data. *The Journal of Negro Education*, 73(4), 408–423.
- Noltemeyer, A. L., & Fenning, P. (2013). Systemic school discipline: Issues of equity from a social justice perspective. In D. Shriberg, S. Y. Song, A. H. Miranda, & K. M. Radliff (Eds.), *School psychology and social justice: Conceptual foundations and tools for practice*. (pp. 91–117). New York, NY: Routledge.
- Noltemeyer, A. L., & McLoughlin, C. S. (2010). Changes in exclusionary discipline rates and disciplinary disproportionality over time. *International Journal of Special Education*, 25(1), 59–70.
- Noltemeyer, A. L., & Ward, R. M. (2015). Relationship Between School Suspension and Student Outcomes: A Meta-Analysis. *School Psychology Review*, 44(2), 224–240.
- \*Obsuth, I., Sutherland, A., Cope, A., Pilbeam, L., Murray, A. L., & Eisner, M. (2016). London Education and Inclusion Project (LEIP): Results from a Cluster-Randomized Controlled Trial of an Intervention to Reduce School Exclusion and Antisocial behaviour. *Journal of Youth & Adolescence*. <http://doi.org/10.1007/s10964-016-0468-4>
- \*Okonofua, J. A., Pauneskua, D., & Walton, G. M. (2016). Brief Intervention to Encourage Empathic Discipline Cuts Suspension Rates in Half Among Adolescents. *PNAS*, 113(19), 5221–5226. <http://doi.org/10.1073/pnas.1523698113>
- Okonofua, J. A., Walton, G. M., & Eberhardt, J. L. (2016). A vicious cycle: A social – psychological account of extreme racial disparities in school discipline. *Perspectives on Psychological Science*, 11(3), 381–398. <http://doi.org/10.1177/1745691616635592>
- Oliver, R. M., Wehby, J. H., & Reschly, D. J. (2011). Teacher Classroom Management Practices: Effects on Disruptive or Aggressive Student behaviour. *Campbell Systematic Reviews*, (4), 1–55. <http://doi.org/10.4073/csr.2011.4>
- \*Panayiotopoulos, C., & Kerfoot, M. (2004). A home and school support project for children excluded from primary and first year secondary school. *Child and Adolescent Mental Health*, 9(3), 109–114. <http://doi.org/10.1111/j.1475-3588.2004.00091.x>
- \*Peck, M. (2006). *The effects of group involvement in a prevention program on at-risk youth's grades, attendance, substance use, and contacts with juvenile probation: An archival study of the STOP program*. Alliant International University, Fresno, US. Retrieved from <http://search.proquest.com/docview/304703141?accountid=9851>; [http://tf5lu9ym5n.search.serialssolutions.com/?ctx\\_ver=Z39.88-2004&ctx\\_enc=info:ofi/enc:UTF-8&rft\\_id=info:sid/ProQuest+Dissertations+&+Theses+A&I&rft\\_val\\_fmt=info:ofi/fmt:kev:mtx:dissertation](http://tf5lu9ym5n.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rft_id=info:sid/ProQuest+Dissertations+&+Theses+A&I&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation)
- Petrosino, A., & Soydan, H. (2005). The impact of program developers as evaluators on criminal recidivism: Results from meta-analyses of experimental and quasi-experimental research. *Journal of Experimental Criminology*, 1, 435–450.
- \*Reese, S. C., Murphy, R. J., & Filipczak, J. (1981). Assessment of Multiple behavioural Procedures on

- Academic and Social Classroom behaviour. *Psychology in the Schools*, 18(3), 349–355.  
[http://doi.org/10.1002/1520-6807\(198107\)18:3<349::AID-PITS2310180317>3.0.CO;2-Q](http://doi.org/10.1002/1520-6807(198107)18:3<349::AID-PITS2310180317>3.0.CO;2-Q)
- Roberts, I., & Ker, K. (2016). Cochrane: The unfinished symphony of research synthesis. *BioMed Central*, 5, 1–5. <http://doi.org/10.1186/s13643-016-0290-9>
- Rocque, M., & Paternoster, R. (2011). Understanding the antecedents of the “School-to-Jail” link: The relationship between race and school discipline. *Journal of Criminal Law and Criminology*, 101(2), 633–666.
- Roorda, D. L., Koomen, H. M. Y., & Oort, F. J. (2011). The Influence of Affective Teacher – Student Relationships on Students’ School Engagement and Achievement: A Meta-Analytic Approach. *Review of Educational Research*, 81(4), 493–529. <http://doi.org/10.3102/0034654311421793>
- \*Russell, A. (2007). *Influencing Adaptive Functioning in School-Age Children: Implementation and Program Evaluation of The Coping Power Program*. George Fox University, US. Retrieved from <http://search.proquest.com/docview/305152799?accountid=9851>;  
[http://tf5lu9ym5n.search.serialssolutions.com/?ctx\\_ver=Z39.88-2004&ctx\\_enc=info:ofi/enc:UTF-8&rft\\_id=info:sid/ProQuest+Dissertations+&+Theses+A&rft\\_val\\_fmt=info:ofi/fmt:kev:mtx:dissertation](http://tf5lu9ym5n.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rft_id=info:sid/ProQuest+Dissertations+&+Theses+A&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation)
- Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston New York: Houghton Mifflin Company.
- Sharkey, J. D., & Fenning, P. A. (2012). Rationale for designing school contexts in support of proactive discipline. *Journal of School Violence*, 11(2), 95–104. <http://doi.org/10.1080/15388220.2012.646641>
- Sherman, L. W. (2010). Defiance , Compliance and Consilience : A General Theory of, 1–32.
- Sherman, L. W., Farrington, D., Welsh, B. C., & Mackenzie, D. L. (2002). *Evidence-Based Crime Prevention. Revised edition*. New York NY: Routledge.
- \*Shetgiri, R., Kataoka, S., Lin, H., & Flores, G. (2011). A randomized, controlled trial of a school-based intervention to reduce violence and substance use in predominantly Latino high school students. *Journal of the National Medical Association*, 103(9), 932–940.
- Singal, A. G., Higgins, P. D. R., & Waljee, A. K. (2014). Clinical/narrative review. A Primer on effectiveness and efficacy trials. *Clinical and Translational Gastroenterology*, 5(e45), 1–4. <http://doi.org/10.1038/ctg.2013.13>
- Skiba, R. ., Trachok, M., Chung, C., Baker, T., & Hughes, R. (2012). Parsing disciplinary disproportionality: Contributions of behaviour, student, and school characteristics to suspension and expulsion. *Paper Presented at the Annual Meeting of the American Educational Research Association, Vancouver, British Columbia, Canada.*, 1–32.
- Skiba, R., Horner, R. H., Chung, C.-G., Rausch, M. K., May, S. L., & Tobin, T. (2011). Race Is Not Neutral: A National Investigation of African American and Latino Disproportionality in School Discipline. *School Psychology Review*, 40(1), 85–107. Retrieved from <http://search.proquest.com/docview/860230091?accountid=9851>;  
[http://tf5lu9ym5n.search.serialssolutions.com/?ctx\\_ver=Z39.88-2004&ctx\\_enc=info:ofi/enc:UTF-8&rft\\_id=info:sid/ProQuest+education&rft\\_val\\_fmt=info:ofi/fmt:kev:mtx:journal&rft.genre=article&rft.jtit](http://tf5lu9ym5n.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rft_id=info:sid/ProQuest+education&rft_val_fmt=info:ofi/fmt:kev:mtx:journal&rft.genre=article&rft.jtit)
- Skiba, R. J., Chung, C.-G., Trachok, M., Baker, T., Sheya, A., & Hughes, R. (2015). Where should we intervene? In D. J. Losen (Ed.), *Closing the school discipline gap: Equitable remedies for excessive exclusion* (pp. 132–146). New York, NY: Teachers College, Columbia University. P.O. Box 103, 525 West 120th Street, New York, NY 10027.
- Skiba, R. J., Michael, R. S., Nardo, A. C., & Peterson, R. L. (2002). The color of discipline: Sources of racial and gender disproportionality in school punishment. *The Urban Review*, 34(4), 317–342.
- \*Smith, M. E. (2004). *A Mixed Paradigm: Study of a Transformational Learning Program for At-Risk High School Students*. Texas A&M University - Corpus Christi. Retrieved from <http://search.proquest.com/docview/305051636?accountid=9851>;  
[http://tf5lu9ym5n.search.serialssolutions.com/?ctx\\_ver=Z39.88-2004&ctx\\_enc=info:ofi/enc:UTF-8&rft\\_id=info:sid/ProQuest+Dissertations+&+Theses+A&rft\\_val\\_fmt=info:ofi/fmt:kev:mtx:dissertation](http://tf5lu9ym5n.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rft_id=info:sid/ProQuest+Dissertations+&+Theses+A&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation)
- \*Snyder, F. J., Flay, B. R., Vuchinich, S., Acock, A., Washburn, I. J., Beets, M. W., & Li, K. K. (2010). Impact of a social-emotional and character development program on school-level indicators of academic achievement, absenteeism, and disciplinary outcomes: A matched-pair, cluster randomized, controlled trial. *Journal of Research on Educational Effectiveness*, 3(1), 26–55.
- Spier, E., Britto, P., Pigott, T., Kidron, Y., Roehlkepartain, E., Scales, P., ... Glover, J. (2013). Parental , Familial , and Community Support Interventions to Improve Children’s Literacy in Developing Countries : A Systematic Review, (July), 1–11.
- Spink, L. A. (2011). Disciplinary exclusion: The impact of intervention and influence of school ethos. *School of Education, Communication and Language Sciences. Doctorate Thesis*, (August), i-80.
- \*Sprague, J. ., Biglan, A., Rusby, J., Gau, J., & Vincent, C. (2016). *Implementing School wide PBIS in Middle Schools: Results of a Randomized Trial*.

- Sulger, P., Sutherland, A., & Eisner, M. (2013). The treatment effect of school exclusion: A propensity score matching approach.
- Sullivan, A. L., Van Norman, E. R., & Klingbeil, D. A. (2014). Exclusionary Discipline of Students With Disabilities: Student and School Characteristics Predicting Suspension. *Remedial and Special Education*, 35(4), 199–210. <http://doi.org/10.1177/0741932513519825>
- \*Tilghman, W. S. (1988). *The effect of peer counseling on lowering the rate of middle school suspensions*. Florida Atlantic University. Retrieved from <http://search.proquest.com/docview/303587508?accountid=9851>; [http://tf5lu9ym5n.search.serialssolutions.com/?ctx\\_ver=Z39.88-2004&ctx\\_enc=info:ofi/enc:UTF-8&rft\\_id=info:sid/ProQuest+Dissertations+&+Theses+A&I&rft\\_val\\_fmt=info:ofi/fmt:kev:mtx:dissertation](http://tf5lu9ym5n.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rft_id=info:sid/ProQuest+Dissertations+&+Theses+A&I&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation)
- Tolan, P., Henry, D., Schoeny, M., & Bass, A. (2008). Mentoring Interventions to Affect Juvenile Delinquency and Associated Problems. *Campbell Systematic Reviews*, (16), 1–112. <http://doi.org/10.4073/csr.2008.16>
- Tyler, T. R., & Huo, Y. J. (2002). *Trust in the law. Encouraging public cooperation with the police and courts*. New York: Russell Sage Foundation.
- U.S. Department of Education. (2012). School Climate and Discipline: Know the Data. Retrieved from <http://www2.ed.gov/policy/gen/guid/school-discipline/data.html#civil-rights-data>
- Valdebenito, S., Eisner, M. P., Farrington, D. P., Ttofi, M. M., & Sutherland, A. (2018). School-based interventions for reducing disciplinary school exclusion: A systematic review. *Campbell Systematic Reviews*, (1). <http://doi.org/DOI: 10.4073/csr.2018.1>
- Van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M., & Holland, R. W. (2010). The Implicit Prejudiced Attitudes of Teachers: Relations to Teacher Expectations and the Ethnic Achievement Gap. *American Educational Research Journal*, 47(2), 497–527. doi:10.3102/0002831209353594
- \*Ward, B., & Gersten, R. (2013). A Randomized Evaluation of the Safe and Civil Schools Model for Positive behavioural Interventions and Supports at Elementary Schools in a Large Urban School District. *School Psychology Review*, 42(3), 317–333.
- Way, S. M. (2011). School discipline and disruptive classroom behaviour: The moderating effects of student perceptions. *The Sociological Quarterly*, 52(3), 346–75. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/22081797>
- Wilson, S. J., Lipsey, M. W., & Soydan, H. (2003). Are mainstream programs for juvenile delinquency less effective with minority youth than majority youth? A meta-analysis of outcomes research. *Research on Social Work Practice*, 13(1), 3–26. <http://doi.org/10.1177/1049731502238754>
- \*Wyman, P., Cross, W., Brown, C., Yu, Q., Tu, X., & Eberly, S. (2010). Intervention to Strengthen Emotional Self-Regulation in Children with Emerging Mental Health Problems: Proximal Impact on School behaviour. *Journal of Abnormal Child Psychology*, 38(5), 707–720. <http://doi.org/10.1007/s10802-010-9398-x>

## TABLES

Table 1. Electronic searches

Databases	
1.	Australian Education Index (AEI)
2.	British Education Index (BEI)
3.	The Campbell Collaboration Social, Psychological, Educational and Criminological Trials Register (C2-SPECTR)
4.	BMJ controlled trials
5.	CBCA Education (Canada)
6.	ClinicalTrial.gov
7.	Criminal Justice Abstracts
8.	Cochrane Central Register of Controlled Trials (CENTRAL)
9.	Database of Abstracts of Reviews of Effects (DARE)
10.	Dissertation Abstracts
11.	Educational Resources Information Centre (ERIC)
12.	EThOS (Beta)
13.	EMBASE
14.	Google
15.	Google Scholar
16.	Index to Theses Database
17.	Institute of Education Sciences - What Works Clearinghouse
18.	ISI Web of Knowledge
19.	MEDLINE
20.	The National Dropout Prevention Centre/Network
21.	The Netherlands National Trial Register (NTR)
22.	Open Grey
23.	Psych INFO
24.	Sociological Abstracts
25.	Social Sciences Citation Index (SSCI)
26.	Scientific Electronic Library Online (SciELO). Electronic database collecting scientific production from developing countries (Spanish and Portuguese)
27.	World Health Organisation International Clinical Trials Registry (WHO ICTRP)

Table 2. Key words for searches

Type of study	Interventions	Population	Outcomes
Evaluation	Disciplinary methods	Schoolchildren	School exclusion
Effectiveness	Token economy	Pupils	Suspension
Intervention	Classroom management program/	Children	Out-of-school suspension
Programme	intervention/ strategies	Adolescents	In-school suspension
Programme effectiveness	School management	School-aged children	Out-of-school exclusion
Impact	Early interventions	Student	In-school exclusion
Effect	School support projects	Youth	Suspended
Experimental evaluation	Skills training	Adolescents	Expelled
Quasi-experimental evaluation		Young people	Expulsion
RCT			Outdoor suspension
Random evaluation			Stand-down
Efficacy trial			Exclusionary discipline
			Discipline

Table 3. Synthesis of the reasons for the exclusion of 471 papers

Reason for exclusion	<i>k</i>	%
Outcome measure was absent	52	11.0
Type of intervention	53	11.2
Methodological design	339	72.0
Participants	5	1.1
Time span	5	1.1
Pilot study	1	0.2
Reports based on the same data	11	2.3
Not enough data for meta-calculations	5	1.1
Total	471	100

Table 4. Summary of data extracted from each included report

	Sample	School-based Programme	Length of intervention <sup>n</sup>	Mean <sup>1</sup> Age (SD)	Type of study	Conflict of interest	Cluster <sup>2</sup>	SMD	CI lower limit	CI upper limit	p-value
Allen et al. 1997	695	Teen outreach	35 weeks	15.8 (1.13)	Journal article	Unclear	No	0.34	0.19	0.50	0.00
Atter, 2005	52	Positive Alternative Learning Support (PALS)	18 weeks	Secondary school	Journal article	Possible	No	0.00	-0.63	0.63	1.00
Barnes et al. 2003	45	Stress reduction	12	16 (1.3)	Journal article	Unlikely	No	0.91	0.47	1.36	0.00
Berlanga 2004	80	Grades, Attendance and Behaviour (GAB)	12	Eighth grade	PhD Thesis	Unlikely	No	0.32	-0.03	0.67	0.08
Bradshaw et al. 2012	12334	School-Wide Positive Behavioural Interventions and Support	4 years	Elementary school	Journal article	Unlikely	Yes	-0.17	-0.29	-0.05	0.00
Bragdon 2010	68	Teach Team Project	9	--	PhD Thesis	Possible	No	0.06	-0.29	0.40	0.74
Brett 1993	126	Efficacy, DC	12	12-14 years	PhD Thesis	Unlikely	Yes	0.09	-0.43	0.61	0.73
Burcham 2002	71	Social problem solving skills training	30	Middle school	PhD Thesis	Possible	No	0.05	-0.29	0.38	0.78
Cook et al. 2014	106	Pro-social skills training	27	5-14 years	PhD Thesis	Unclear	No	0.35	-0.19	0.90	0.20
Collier 2002	60	BAM (skills-training) and MATCH (tutoring)	8	--	Technical report	Possible	No	1.10	0.72	1.48	0.00
Cornell et al. 2012	201	Threat assessment	--	--	Journal article	Possible	Yes	0.58	0.25	0.91	0.00
Crowder 2001	109	Gang Resistance, Education and Training (GREAT)	9	--	PhD Thesis	Unlikely	No	0.07	-0.31	0.45	0.72
Dynarski et al. 2003	968	21 <sup>st</sup> Century Community Learning	1 year	Elementary school	Technical report	Unlikely	No	0.20	-0.13	0.53	0.24
Edmunds et al. 2012	1607	Early College High School	--	15.3	Journal article	Unlikely	No	0.44	0.24	0.64	0.00
Farrell et al. 2001	626	Academic skills enhancing Responding in Peaceful and Positive Ways (RIPP)	25	11.7 (0.6)	Journal article	Possible	Yes	0.42	-0.13	0.96	0.13
Feindler 1984	36	Anger control training	7	13.8 (.68)	Journal article	Unclear	No	1.20	0.72	1.68	0.00
Harding 2011	48	Over to you	6	Eighth grade	PhD Thesis	Possible	No	0.09	-1.15	1.33	0.89
Hawkins et al. 1988	160	Proactive Classroom Management	1 year	Seventh grade	Journal article	Unlikely	No	0.23	-0.10	0.56	0.17
Hirsch et al. 2011	535	After School Matters	20	15.9	Technical report	Unlikely	No	0.02	-0.12	0.16	0.75
Hostetter & Fisher 1997	317	Project CARE (Skill for	1 year	Third grade	Journal article	Unlikely	No	0.20	0.03	0.36	0.02

<sup>1</sup> When the mean age was not available in the original study, grade at school has been reported. Grade at school allow the reader to have a general idea of the age of students.

<sup>2</sup> Due to the nature of the settings (schools) some studies reported clustered data. We corrected SE errors when it was needed.





Table 5. Sub-group analyses

Moderators	SMD	95% CI	p-value	k	Groups comparison
<b>Gender of school population</b>					
Predominantly masculine	.41	(.10, .72)	p<.05	11	X <sup>2</sup> =1.84; df=1; p>.05
Mixed gender	.17	(.02, .32)		19	
<b>Age</b>					
Elementary school	.27	(.09, .45)	p<.05	12	X <sup>2</sup> =1.81; df=2; p>.05
Middle school	.23	(.04, .41)	p<.05	16	
High school	.45	(.18, .72)	p<.001	8	
<b>Theoretical bases of the intervention</b>					
Individual oriented	.33	(.02, .93)	p<.05	28	X <sup>2</sup> =.48; df=1; p>.05
School oriented	.25	(.04, .45)	p<.05	10	
<b>Training before implementation</b>					
Reported	.29	(.16, .43)	p<.001	26	X <sup>2</sup> =.16; df=1; p>.05
No reported	.34	(.15, .53)	p<.001	12	
<b>Monitoring during intervention</b>					
Reported	.20	(.05, .35)	p<.05	15	X <sup>2</sup> =2.89; df=1; p>.05
No reported	.37	(.25, .50)	p<.001	23	
<b>Reasons for conducting research</b>					
Demonstration	.43	(.26, .59)	p<.001	16	X <sup>2</sup> =8.15; df=1; p<.05
Routine	.13	(.00, .25)	p<.05	18	
<b>Cluster versus individual level studies</b>					
Cluster randomized units	.22	(.02, .41)	p<.05	13	X <sup>2</sup> =1.38; df=1; p>.05
Individually randomized units	.36	(.22, .50)	p<.001	25	
<b>Evaluator role</b>					
Independent	.13	(.00, .25)	p<.05	18	X <sup>2</sup> =12.36; df=1; p<.001
Dependent	.47	(.32, .62)	p<.001	16	
<b>Type of intervention</b>					
Enhancement of academic skills	.43	(.25, .61)	p<.001	2	X <sup>2</sup> =18.4; df= 8; p<.05
After-school programme	.05	(-.08, .17)	p>.05	2	
Mentoring/monitoring	.47	(.02, .93)	p<.05	5	
Skills training for students	.31	(-.05, .67)	p>.05	9	
Skills training for teachers	.31	(.11, .52)	p<.05	4	
School-wide strategies	.20	(-.03, .43)	p>.05	6	
Violence reduction	.48	(-.33, 1.3)	p>.05	3	
Counselling, mental health	.46	(.23, .68)	p<.001	3	
Other	.21	(.03, .39)	p<.05	4	

Table 6. Meta-regression results: a priori moderators

Predictors	Model I			Model II		
	b	SE	95%CI	b	SE	95%CI
Intercept	.37	.33	-.28, 1.06	.37	.34	-.30, 1.04
Participants' characteristics						
Gender						
Male predominant	.09	.38	-.64, .82	.11	.39	-.66, .87
Mixed	-.05	.37	-.77, .67	-.06	.38	-.81, .69
Grade at school (High)						
Elementary	-.11	.21	-.52, .29	-.16	.23	-.61, .29
Middle	-.22	.20	-.60, .16	-.24	.20	-.64, .16
Mixture	.17	.40	-.62, .96	.04	.48	-.90, .98
% of White	-.00	.004	-.008, .006	-.00	.003	-.00, .00
Intervention characteristics						
Individual versus school level change				.11	.21	-.30, .52

\*  $p < .05$

Table 7. Meta-regression results: post-hoc moderators

Post-hoc predictors	Model b	SE	95%CI
Intercept	.32	.17	-.01, .07
Characteristics of the research			
The ends of the research (routine vs. demonstration)	.06	.14	-.20, .33
Role of the evaluator (independent vs. dependent)	-.36	.14	-.63, -.09*
Risk of quality bias	.008	.01	-.01, .03

\*  $p < .05$

Table 8. Sensitivity analysis

Panel A: with outlier							
Covariate	SMD	95% CI	SE	V	n	k	Measure of heterogeneity
0.50	0.31	(.20, .43)	0.058	0.003	36	37	$Q=259.0$ ; $df=36$ ; $p<.001$ ; $I^2=86$ ; $\tau^2=.081$
0.75	0.32	(.21, .44)	0.058	0.003	37	38	$Q=338.4$ ; $df=37$ ; $p<.001$ ; $I^2=89$ ; $\tau^2=.090$
Panel B: without outlier							
0.50	0.27	(.16, .38)	0.056	0.003	36	37	$Q=233.3$ ; $df=36$ ; $p<.001$ ; $I^2=84.5$ ; $\tau^2=.072$
0.75	0.28	(.17, .39)	0.055	0.003	36	37	$Q=288.0$ ; $df=36$ ; $p<.001$ ; $I^2=87.5$ ; $\tau^2=.076$
Panel C: with outlier Winsorised							
0.50	0.30	(.19, .41)	0.057	0.003	37	38	$Q=246.23$ ; $df=37$ ; $p<.001$ ; $I^2=84.9$ ; $\tau^2=.076$
0.75	0.30	(.19, .41)	0.055	0.003	37	38	$Q=301.3$ ; $df=37$ ; $p<.001$ ; $I^2=87.7$ ; $\tau^2=.078$